

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 15:14:14 ; Search time 1962.38 Seconds
(without alignments)
11547.332 Million cell updates/sec

Title: US-09-484-964-1
Perfect score: 1465
Sequence: 1 cgggaagattgttaaccc.....gataaatggcagccagcc 1465

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_bal: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pi1: *
13: gb_pi2: *
14: gb_pi3: *
15: gb_pi4: *
16: em_bal: *
17: em_ba2: *
18: em_fun: *
19: em_hcgo_hum: *
20: em_hcgo_inv: *
21: em_hcgo_rod: *
22: em_hcgo_hum1: *
23: em_hcgo_hum2: *
24: em_hcgo_hum3: *
25: em_hcgo_hum4: *
26: em_hcgo_hum5: *
27: em_hcgo_hum6: *
28: em_hcgo_hum7: *
29: em_hcgo_hum8: *
30: em_hcgo_inv1: *
31: em_hcgo_inv2: *
32: em_hcgo_other: *
33: em_hcgo_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pi: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_v1: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_v11: *
59: gb_v12: *
60: gb_hcgo1: *
61: gb_hcgo2: *
62: gb_hcgo3: *
63: gb_hcgo4: *
64: gb_hcgo5: *
65: gb_hcgo6: *
66: gb_hcgo7: *
67: gb_hcgo8: *
68: gb_hcgo9: *
69: gb_hcgo10: *
70: gb_hcgo11: *
71: gb_hcgo12: *
72: gb_hcgo13: *
73: gb_hcgo14: *
74: gb_hcgo15: *
75: gb_hcgo16: *
76: gb_hcgo17: *
77: gb_hcgo18: *
78: gb_hcgo19: *
79: gb_hcgo20: *
80: gb_hcgo21: *
81: gb_hcgo22: *
82: gb_hcgo23: *
83: gb_hcgo24: *
84: gb_hcgo25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_pr10: *
95: gb_pr11: *
96: gb_pr12: *
97: gb_pr13: *
98: em_ba3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	1286.6	87.8	1514 9	AR087051
2	1286.6	87.8	1514 97	HSU083117
3	1120.2	76.5	1223 97	HSU061397
4	1120.2	76.5	1227 91	BC006462
5	1079	73.7	160939 76	AC079354
6	1069.2	73.0	155913 81	AL513282
7	1069.2	73.0	167440 80	AL559981
8	1000.2	68.3	1017 97	HSU08784

9	813	55.5	163196	85	AC005220
10	770	52.6	816	97	HS067122
11	765.8	52.3	1187	94	AF033353
12	743.6	50.8	235411	76	AC084066
13	744.4	49.4	138519	89	AL135926
14	614.4	41.9	106168	87	AC011450
15	614.4	41.9	178733	73	AC068786
16	574.2	39.2	162776	66	AC021171
17	537.8	36.7	146267	76	AC008815
18	537.8	36.7	160970	61	AC010501
19	537.8	36.7	173048	87	AC008653
20	529	36.1	124764	60	AC008377
21	529	36.1	209022	71	AC034212
22	527.8	36.0	163082	67	AC022142
23	520.8	35.5	161190	67	AC016567
24	432.2	29.5	249997	60	AC009042
25	420.2	28.7	590	93	HSSMT3C
26	403	27.5	510	7	AF242526
27	390	26.6	151977	63	AC015615
28	390	26.6	159252	83	AP003168
29	390	26.6	160452	71	AC040946
30	372.4	25.4	180821	68	AC023974
31	370.8	25.3	205741	60	AC008548
32	368.8	25.2	372	9	AX021844
33	304.4	20.8	306	97	HS072722
34	287	19.6	82688	61	AC010493
35	253.4	17.3	310360	75	AC074172
36	247	16.9	310360	85	AC074172
37	215.2	14.7	466	8	XLSTMO
38	191.6	13.1	105692	67	AL145106
39	191.6	13.1	135623	91	AC022618
40	179.8	12.3	670	8	AB036430
41	115.4	7.9	193	54	G24133
42	103.8	7.1	212093	64	AC015909
43	98.8	6.7	191	92	HS16F8F
44	97.8	6.7	295	92	HS16F8F
45	97.4	6.6	181861	72	AC064836

ALIGNMENTS

RESULT 1
AR087051 1514 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 1 from patent US 5985664.
ACCESSION AR087051
VERSION AR087051.1 GI:10013817
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1514)
AUTHORS Baker,B.F. and Cowser,L.M.
TITLE Antisense modulation of Sentrin expression
JOURNAL Patent: US 5985664-A 1 16-NOV-1999;
FEATURES
source Location/Qualifiers
1..1514

BASE COUNT 462 a 256 c 314 g 482 t
ORIGIN
Query Match 87.8% Score 1286.6; DB 9; Length 1514;
Best Local Similarity 94.1%; Pred. No. 9.1e-239;
Matches 1384; Conservative 0; Mismatches 74; Indels 13; Gaps 4;

QY 2 ggaaggaatttaaacccggagagaggttctgctcccgagcgctgctgagga 61
DB 50 gggaaggaattttaaaccggagagaggttctgctcccgagcgctgctgagga 109
QY 62 gaccccggtgaagcaccgctcattcattctgacaggaagcacaactcactgag 121

DB 110 GACCCCGGGTGAAGCCACCGTCATCTGTGACGAGGACCAAACTTCACGTGAG 169
QY 122 acttgggggataaagcaagatgaatataatataactcaagtcattggagagataca 181
DB 170 ACTTGGGGGATTAAGAGAGAGGATATATTAACTCAAAATGCTATGGACAGATGCA 229
QY 182 gtagatcacttcaagtgaatgaacacacatcacaagaactcaagaatcact 241
DB 230 GAGAGATTCATTAAG 289
QY 242 gtaacagaagaggttccatgaatcactcaggttctccttgaaggtcaagaatg 301
DB 290 GTCAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 349
QY 302 ctgataatcatactccaaaagacacaggaatggaaggaagatgtaagttatc 361
DB 350 CAGAT 409
QY 362 aggaacaaagggggggtcattcaacagtttagatctccttcttcttcttcc 421
DB 410 AGGAAACAAAGGGGGGATCAATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 469
QY 422 tcaatccttttattttaaataagttccttcttcttcttcttcttcttcttcttcc 481
DB 470 TCAATCCTTTTATTTTAAATAGTTCTTTTGTATGTCGTTTCAAAACGGAATTG 529
QY 482 aaactggaacccatccttcttgaacacatcgtgaatttgaattctagtcattc 541
DB 530 AAACCTGGACCCCATCTCTTGAACATCTGTAATTTGAATTTGATGCTATATTC 589
QY 542 attatgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 601
DB 590 ATTATGTTTGTTCATATGTCGATTTTGTGATCAACCTCAAGCTCCCTCATATT 649
QY 602 accctccctttaaataatcagctgtgacagagaggtcccttcttcaagacatga 661
DB 650 ACCCTCTCTTTTAAATTAAGGTCACAGAGAGGTCACCTTTTCAAGACATTTGA 709
QY 662 tttaagcttggtgtgtaataagatcgacacatgcaagtgcttaataagcttcca 721
DB 710 TTTTGGGCTTTGTTCTTGA-----TGTGATTTACTTCACTCCGAGCTGTGACTTTTCAG 764
QY 722 atggcccttgatgtcagcagatgattactcactcctcctcctcctcctcctcctc 781
DB 765 TGGGAGATGAAGTTTGTGTGATATAATAGATCAACCAATGCAAGTGTCTTAATGA 824
QY 782 atggaattt-----tcagagaactggaactggaagaaatgaccttcttaactga 835
DB 825 CTTTCAATTTGGCCCTGACAGAGAACTGAACCTGGAATAATGACCTTTCCTTAATGGA 884
QY 836 agctactttaaata-tttgaggggtctgacacaaagaaaggaatatacaggttgaagta 894
DB 885 AGCTACTTTTAAATTTTGGAGGCTGTGACCAAAAGAGAGGATATCAAGTGAAGTCA 944
QY 895 agatgacagataaggtgagatgaatgactaactccaaagatggtcctcaagaaagag 954
DB 945 AGATGACAGATTAAGTGAAGTATGACTAATCTCAAAAGATGCTTCACTGAGAAAGG 1004
QY 955 catttaagatttttaaataatctgtcagaagaatcccaagaagttctaatttcaat 1014
DB 1005 CATTTAAGATTTTAAATATCTGTGACAGAGATCCCAAAAGTTCTTAATTTCAATTA 1064
QY 1015 gcaatlaataagctatatactcagaataatatacacaagaacagccttttgatt 1074
DB 1065 GCATTTAATAAGTATATATGCAAGAAATATAATCAACAGAACACTGCTTTT-TGATT 1123
QY 1075 ttattgtacttttggcctggagataggttttaaatgacatgtgctgacagcttc 1134
DB 1124 TTATTGTACTTTTGGCCTGGGATATGGGTTTAAATGACATGTGTCACACACTTC 1183
QY 1135 attaaataaacaatttttcaaaaatgtaataatgcttaatttatttaattatgtata 1194
DB 1184 ATTAAATTAACAAATATTTGTCAAAATGCTACTAATGCTTATTTTATTTATTTGATA 1243

Db	1184	ATTAAATAAACATATTGTGCATAAAATCGTACTANTGCTATTATTATTATTATTGTATA	1243
QY	1195	gaagaaaaaaatgcctaaaaaagaagtttcctgcataatactgynaatgacatggt	1254
Db	1244	GAAGAAAAAAATTCCTTAATAATAGGTTTTCTTCGATAAATACTGGAAATTCACATGCT	1303
QY	1255	aaaatttttcttcattactgacagagatgtaagttaagtccttttgagacatgaagt	1314
Db	1304	ACAATTTTTTCTTCATTACTGTCACAGGATGATGTTTAATGACTTTGGAGCACATGAAGT	1363
QY	1315	tactgaagtcgcttcgaaatcaaggaatttaataagggacaaatcactttaaatacca	1374
Db	1364	TACTGAAGTGCCTTCGAATCAAGGATTTAATTAAGGCCACAATACCTTTTAATACCA	1423
QY	1375	gtgttcggtttttttaaaacttgatatcccgatagtgtgcatttgatacaggtacc	1434
Db	1424	GTGTTCGTTTTTTTAAAACTTGATATTCOCGTGTGTCATATTGATACAGGTACC	1483
QY	1435	caatcactgtgataaattggcactgcacgc	1465
Db	1484	CAATCATGTTGGATAAATGGGCAATGCCACGC	1514

RESULT	3
LOCUS	HSU61397
DEFINITION	HSU61397 1223 bp mRNA
ACCESSION	Human ubiquitin-homology domain protein P1C1
VERSION	061397.1 GI:1518693
KEYWORDS	
SOURCE	human.
	14-JAN-1997
	complete cds

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 1223)
Boddy,M.N., Howe,K., Etkin,L.D., Solomon,E. and Freemont,P.S.
PIC 1, a novel ubiquitin-like protein which interacts with the p

JOURNAL
Oncogene 13 (5), 971-982 (1996)
MEDLINE 8870663-3

REFERENCE	2 (bases 1 to 1223)
AUTHORS	Howe, K., Boddy, M.N., Etkin, D.E., Solomon, E. and Freemont, P.S.
TITLE	Direct Submission
JOURNAL	Submitted (20-JUN-1996) Somatic Cell Genetics, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK
FEATURES	Location/Qualifiers
source	1. .1223

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="placenta"
146. .451
CDS

```

BASE COUNT	371 a	209 c	268 g	375 t
ORIGIN	/product="ubiquitin-homology domain protein PIC1" /protein_id="AAB40388.1" /db_xref="GI:1518694" /translation="MSDQEAKPSTEDLGKKEGEYTKLKVTIGQDSSLEHFFVKMTTHL KKLVESTQQRGVPMNSLRFLLFEGQRIADNHPTPKELGMEEDVEIVYEQTIGGHSY"			

Query Match	76.5%;	Score 1120.2;	DB 97;	Length 1223;
Best Local Similarity	99.5%;	Pred. No. 1.2e-206;		
Matches 1155; Conservative	0;	Mismatches 3;	Indels 3;	Gaps 3

[illegible]

Db	120	GACCCCGGGGTAAAGCCACCGTCATCAGTCTGACAGGAGGCAAAACCTTCAACTGAGG	179
QY	122	acttgggggacaagaagaaaggatgaaatataatbaactcaagtcattggacaagataga	181
Db	180	ACTTGGGGGATTAAGAAAGGAGGTGAATATTAATTAACCTCAAGATCATTTGGACAGATAGCA	239
QY	182	gtagaattacttcaaaagtgaaaaatgaaacaacactctaaaaaactcaagaatcactact	241
Db	240	gtacagtttacttccaaagtcgaaatgaaatgacacacactctcaaaaactcaaaaactcaact	299
QY	242	gtcaaaagcaagggtgtcccaatgaaattactcagggttctctcttggaggtcaagaaatg	301
Db	300	gtcAAAGACAGGGGTGTCCAAATGAATTCACACAGGTTCCTCTTGAAGGTCAAGAAATG	359
QY	302	ctgataatcatatctcccaaaagacacgggaatgagagaaagaatgtgattgaattatc	361
Db	360	CTGATATATATCTCCAAAAGAACTGGGAATGGAGGAAGAAAGATGATTAAGATTATC	419
QY	362	aggaaacaaacggggggtcattcaacaagttatgatctcttcttcttcttcttcc	421
Db	420	AGGAACAAACGGGGGTCATTCACAGTTTGGATTTCTTTATTTTTTCTCTTCC	479
QY	422	tcaatccttttatttctttaaanaagtccttcttgaatgtygtgttcaaaagyaatg	481
Db	480	TCAATCCCTTTTATTTTAAAAAATAGTCTTTTGTATGGTGGTTCAAAAGGAAATG	539
QY	482	aaaactggcaccccactctcttggaaaactcgtgaatttgaatctgagtctcaatc	541
Db	540	AAAACGTGCACCCCATCTCTTTGAACATCTGGTAATTTGAATTTAGTCTCATTAATTC	599
QY	542	attattgttgtttcattcattgtctcgaatttttggtagtcaaacgcctcagtcacctat	601
Db	600	ATTATTGTGTGTTTCATATGTGTCATTTTGGTATCAAGCCATGCCATGCCCTTATAT	659
QY	602	accctctcctttttaaanaatgaatgtgacaaagagagtgaccttttcaagacattgca	661
Db	660	ACCTCTCTCTTTTAAAAATTAATTAACGTGTGCACAGAGAGTCACTTTTTCAGCAATTGCA	719
QY	662	ttttcaagcttgttgatgataaataagacacgaatbcaagtgattcaataatgacttcca	721
Db	720	TTTTGAGCTTGTGTGTGAATTAATTAATGATGACCAATGCAAGTGTTCATATGACTTTCCA	779
QY	722	attggccctgtatgtctagaatgtgatcttaactcctctggactgtgacttcaagtgg	779
Db	780	ATTGGCCCTGTATGTTCAGCATGTGATTACTTCACTCTGGACTGTGACTTTCAATGGG	839
QY	780	agatgaaagttttccaagaactbaactctgagaaaatgaccttcttaacttaactgaagt	839
Db	840	AGATGGAAGTTTTCAGAGAAGTGAATGTGGAAAAATGACTCTTCTTAACTTGAAGCT	899
QY	840	acctttaaatttgaagggtctggaacaaagaagagaaatcgaattgaagtgaatg	899
Db	900	ACTTTTAAAAATTTGAGGCTGTGACCAAAAAGAGAGAAATATACGTTGAATCTCAAGATG	959
QY	900	acagataaagtgtgagaaatgaacttaactccaagaatgtggttcaacttgaagaagaacatt	959
Db	960	ACAATATAAGGAGAGATTAATGACTTAATCCAAAAGATGGCTCTCACTGAAMAAAAAGCATTT	1019
QY	960	taagattttttaaanaactctgtcagaagaalccccaagaagttctaatttcaatgacat	1019
Db	1020	TAAATTTTAAAAAATCTTGTGAGAAAGATCCAGAAAAGTTCTAATTTTCAATTGCAAT	1079
QY	1020	taataaagctatcatgacgaaatgaaatacacaagaacacgtcctcttlttgatttatt	1079
Db	1080	TAAATAAAGCTATACATGACCAAAATGAATACAAACAGAACCTCTCTTTTTCATTTTAAT	1138
QY	1080	tgtacttttggccttggatattgtgtttaaagttaaatgacattgtctgtacaaacttcaaa	1139
Db	1139	TGTACTTTTGGCCTGGGATATGAGTTTAATTAATGACATGTGCTTACAGCTTCATTTAA	1198
QY	1140	aataaacaatattgtccaana	1160
Db	1199	AATTAACAAATATTGTAAAAA	1219

Wegman, 721-250 74601762

RESULT 4
 BC006462 1227 bp mRNA PRI 13-APR-2001
 LOCUS Homo sapiens, ubiqutin-like 1 (sentrin), clone MGC:2095, mRNA,
 DEFINITION complete cds.
 BC006462
 ACCESSION BC006462.1 GI:13623670
 VERSION MGC.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 1227)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 Info@bcsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McEavey, Steven
 Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 8 Row: f Column: 12.
 FEATURES
 source
 location/Qualifiers
 1. 1227
 /organism="Homo sapiens"
 /db_xref="LocusID:7341"
 /db_xref="taxon:9606"
 /clone="MGC:2095 IMAGE:3546125"
 /tissue_type="Placenta, chorioncarcinoma"
 /clone_lib="NIH-MGC-21"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 132..437
 /codon_start=1
 /product="ubiquitin-like 1 (sentrin)"
 /protein_id="AAH06462.1"
 /db_xref="GI:13623671"
 /translation="MSDQKPKSTEDGDKKEGYIKLVYIGDSSIEHFKVMTHL
 KRIKESYCGROGVPNMSIRLFEGRADNHTPEKELMEDEYIEVYQEDTGHSTV"
 BASE COUNT 386 a 208 c 259 g 374 t
 ORIGIN
 Query Match 76.5%; Score 1120.2; DB 91; Length 1227;
 Best Local Similarity 99.5%; Pred. No. 1.2e-206;
 Matches 1155; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
 Oy 2 gggaaggttgtaaacccggagaggttcttcttccagagccgctgtgcsa 61
 Db 46 GAGAAGATTGTAAACCCGGAGGAGGTCTGTCTACCCGAGCCGCTGTGCGGA 105
 Oy 62 gaccccggtgaagccacgctcatcgtctgacagaggaacaaccttcaactgag 121

Db 106 GACCCCGGGGTGAAGCCACCGTCATCTGTCTGACCGAGGAGCAAAACCTCAAGTAGG 165
 Oy 122 acttgaggatagaagcaagtgtaaalatalaaactcaaaagtcaattgacagataga 181
 Db 166 ACTTGGGGGATTAAGAGAGAGGATGATATATTAAATCAAAAGTCAATGGACAGATAGCA 225
 Oy 182 gtgagattcaactcaagtgtaaaatgacacactcacaagaactcaagaataact 241
 Db 226 GTGAGATTCACTTCAAGTGAATGACACACATCTCAAGAACTCAAGAAATCAACT 285
 Oy 242 gtcaaacagaggtgttccaatgaatcaactcaagttctcttggaggtcagagaatg 301
 Db 286 GTCAAGACAGAGGTTTCCATTAATGATTCACACAGTTTCTCTTGGAGGTCAGAGATTG 345
 Oy 302 ctgatacatcactccaaagaactgggaatggaaggaagaatgtaattgaaattatc 361
 Db 346 CTGATTAATCATTAATCCAAAGACAGTGGGAATGAGAGAAAGATGATGAAGTTATC 405
 Oy 362 aggaacaagggggtgtcatccaagttagatatcttcta-tttttctcttc 420
 Db 406 AGGAACAAACGGGGGTCATTCACAGTTTAGATATCTTTTATTTTCTTTTCC 465
 Oy 421 ctcaatccttttttaaaaaaagttcttcttgaatgtgtgttcaaacgaatt 480
 Db 466 CTCATCTCTTTTATTTTAAATAATAGTCTTTTGTATGTGGGTCAAAACGAAT 525
 Oy 481 gaaacatgacccacatctcttgaacatctgttaattgaattcagttcattat 540
 Db 526 GAAACTGGACACCCCATCTCTTTGAACATCTGTAAATTAATTAATTAATTAAT 585
 Oy 541 catttgctgttcttcaatgtgtcgtatcttggatcaagccctcactccatcat 600
 Db 586 CATTAATGTTTGTTCATGTGCGATTTTGGATGATCAAGCCATGCTCCCATAT 645
 Oy 601 taacctctctttaaataatcagtgtagacagagaggtcaacctttcaggaattgc 660
 Db 646 TACCTCTCTTTTAAATAATACGTGTGACACAGAGATGACCTTTTACAGCAATTGC 705
 Oy 661 atttcaagctgtgtgttaataatgaatcagacaaatcaagtgtcattgaattcc 720
 Db 706 ATTTTGAAGCTGTGTGTATTAATTAATGATGACCAATGCAAGTTCATTAATGATTTCC 765
 Oy 721 aattgacctgtatc-agaatgtatcttcaactcctgtgactgtactcagtg 779
 Db 766 AATTGGCCCTGATGTTCTGATGATGATTAATTAATTAATTAATTAATTAAT 825
 Oy 780 agatgaagttttcagagaactgaactgttgaataaaatgaacttcccttaactgaact 839
 Db 826 AGATGGAAGTTTTCAGAGACAGTGAAGTGTGAAAAATACCTTTCCTTAATGATGCT 885
 Oy 840 acttttaaaatttgagaggtctgagacaaagaagagaatatacagttgaatgaatg 899
 Db 886 ACTTTTAAATTTTGAGGGTCTGAGACCAAAAGAGAGATATCAGTTGAATCAAGATG 945
 Oy 900 acagataagtgtagtaatgaactcaactccaaagatgcttcaactgaagaagaagcatt 959
 Db 946 ACAGATTAAGTGAAGTAATGACTCACTCAAGAGGCTTCACTGAAGAAAGGATTT 1005
 Oy 960 taagatttttaaaatctgtcagaagaatcccaagaagtttaatttcaactgaact 1019
 Db 1006 TAAGATTTTAAATAATCTTGTGAGAGATCCAGAAAGTTTAAATTTTCAATTAAGCAAT 1065
 Oy 1020 taataagctatacatgcaagaatagaatacaagaacactgctcttttagatttatt 1079
 Db 1066 TAATAAAGCTATATCATGCAAGAAATGAATACAGAGAACTGCTCTTTT-T-GATTTATT 1124
 Oy 1080 tgaacttttgacctggagataatggttttaattgaatgttctgttcaagcttcaat 1139
 Db 1125 TGTACTTTTGGCTGTGGATATGGGTTTAAATGACATTTCTGTACAGCTTCAATTAA 1184
 Oy 1140 aataaacaatattgtcaaaa 1160


```

QY      800  actgacgtggaataatgacaccttcccttaactgaacttaaatgaggtc 859
          |||
Db 138601 ACTGAACTGTGAAAAATGACCTTCTTAAGCTACTTAAATTTGAGGCTC 138542
QY      860  tggacacaaagaagaatatacaaggttgaagtcagaatgacagataaggtgagat 919
          |||
Db 138541 TGGACCAAAAGAAGAGATATACAGTTGAATGACAGATGACAGATGAGTAAAT 138482
QY      920  gactactccaagaatggtctcactgaagaagaagcatttaagtttttaaatctt 979
          |||
Db 139481 GACTACTCTCAAGAATGAGCTTCACTGAAAGAAAGCATTTTAAAGTTTAAATCTT 138422
QY      980  gtccagaagatcccaagaagaagttctaatttcaatgaataataaagctatactgca 1039
          |||
Db 138421 GTCAGAAAGTCCCAAGAAAGTCTAATTTCAATTAGCAATTAATAAGCTATACATGACG 138362
QY      1040  aaatgaatacaacaagaacacgtctcttttgaattatttcttcttgcctggagat 1099
          |||
Db 138361 AAATGAATACACAGAACACGCTCTTTT-T-GATTTTATTTGACTTTTGGCTGGGAT 138303
QY      1100  atgggttttaaatgagcattgtctgtacagcttcaataaatacaaatctgtcaaa 1159
          |||
Db 138302 ATGGGTTTAAATGACATTTGCTGTACAGCTTCAATTAATAACATATTTGT-AAA 138244
QY      1160  aatcgactaatgtctatttatttatttataatgtatagaagaataaataagctaaataag 1219
          |||
Db 138243 AATCATACATAAGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 138184
QY      1220  gttttctgataataatcaggaatgtgcacatgtgtacaattttcttcttcttcttcttct 1279
          |||
Db 138183 GTTTTCTTGATTAATACGTGAAATTCACATGTGTACAAATTTTCTTCATTTACTGTAC 138124
QY      1280  aggaatgatgttaactgtctgtgacacgtgaagaagctgaagctgcttcttctgaatcaag 1339
          |||
Db 138123 AGGATGATGATTTATGACTTTGAAGCACTGAAGTTACTGAAGTCTGATGATCAAG 138064
QY      1340  atttaataagccacaataccttcttaatactcaagtgctgtcttcttcttcttcttcttct 1399
          |||
Db 138063 ATTAAATTAAGGCCACAAATACCTTTTAATACAGTGTCTG-TTTTAAAAAATTG 138005
QY      1400  atattcccgatgtgtatatttgatacaagtaacacatatttggataaataagggcatg 1459
          |||
Db 138004 ATATTCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 137945
QY      1460  cca 1462
          |||
Db 137944 CCA 137942

RESULT 6
AL513282/c 155913 bp DNA HTG 26-FEB-2001
LOCUS Homo sapiens chromosome 1 clone RP11-179G5, ** SEQUENCING IN
DEFINITION PROGRAMS ***, 8 unordered pieces.
ACCESSION AL513282
VERSION AL513282.5 GI:13162044
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 155913)
AUTHORS Pavitt,R.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Feb 26, 2001 this sequence version replaced gi:13161719.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk

```

```

Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA1179G5
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 153015 bases at least Q40
Consensus quality: 153938 bases at least Q30
Consensus quality: 154565 bases at least Q20
Insert size: 155213; sum-of-contigs
Insert size: 164818; 1.0% error; agarose-fp
Quality coverage: 6.89x in Q20 bases; sum-of-contigs Quality
coverage: 7.27x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 20838: contig of 20838 bp in length
20839 20938: gap of 100 bp
20939 50589: contig of 29651 bp in length
50590 50689: gap of 100 bp
50690 81538: contig of 30849 bp in length
81539 81638: gap of 100 bp
81639 89682: contig of 8044 bp in length
89683 89782: gap of 100 bp
89783 108105: contig of 18323 bp in length
108106 108205: gap of 100 bp
108206 149395: contig of 41190 bp in length
149396 149495: gap of 100 bp
149496 153075: contig of 3580 bp in length
153076 153175: gap of 100 bp
153176 155913: contig of 2738 bp in length.
*
FEATURES
Source
1. 155913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-179G5"
/clone_id="RP11-179G5"
/location="Qualifiers"
misc_feature
1..20838
/note="assembly_fragment:02743
fragment_chain:1
clone_end:77
vector_side:left"
misc_feature
20939..50589
/note="assembly_fragment:01236
fragment_chain:1"
misc_feature
50690..81538
/note="assembly_fragment:00478
fragment_chain:2"
misc_feature
81639..89682
/note="assembly_fragment:03283
fragment_chain:2"
misc_feature
89783..108105
/note="assembly_fragment:00942
fragment_chain:3"
misc_feature
108206..149395
/note="assembly_fragment:02750
fragment_chain:3"
misc_feature
149496..153075
/note="assembly_fragment:02486
fragment_chain:4"
misc_feature
153176..155913
/note="assembly_fragment:03083
fragment_chain:4"
BASE COUNT 43830 a 34529 c 35326 g 41527 t 701 others
ORIGIN

```

Query Match	73.0%;	Score 1069.2;	DB 81;	Length 155913;
Best Local Similarity	97.2%;	Pred. No. 9.6e-197;		
Matches 1130;	Conservative	0;	Mismatches 28;	Indels 4;
				Gaps 4;

OY	2	gaaagattctgtcaaacccccggagaggtcttgcttaccggagccgtctgtcgga	61
Db	136441	GAGAAGATTGTGTAACCCCCGGAGTAGGTTCTTGCTTACCGAGGCCCTCTGTGGCGA	136382
OY	62	gaaccccggttgaaagccaccgtcatatgtctgaacagagagcaaaccttcaactag	121
Db	136381	GACCCCGGCTGAAGCCACTGTCACTCATGCTGTGACCGAGAGGCAAAACCTTCAACTGAGG	136322
OY	122	acttgggggataagagcaaggttgatatattataaacctaaagtaatttggaaagatgca	181
Db	136321	ACTTTGGGGGATTAAGAAGGAAGGTGAAATATTATTAACCTAAAGTCAATTGGACAGGATGCA	136262
OY	182	gtgagattcaacttcaaatgtgaaatgacaacacatctaaagaaactcaagacatcact	241
Db	136261	GTCAGATTCACCTTCAAAAGTAAAGTAAAGACACACATCTCAAGAAACTCAAGAAATCATACT	136202
OY	242	gtcaaaagacaagggtgtctccaatgaaatcaacaaagttctctcttttggaggttcaagaa	301
Db	136201	GTCAAAGACAGGGGCGTCCCAATGAATTCATTCAGGTTCTCTTTGAGGGTACGAGAAATG	136142
OY	302	ctgataatacatcatcccaaaaagaactcggaatgaggaagaaagatgtatgaaattatc	361
Db	136141	CTGATATATATATCTCCAAAAGTAAAGTAAAGGAGGAAGGAAGATGATTAAGAACTTATC	136082
OY	362	aggaaacaaacggggggtcatctcaacaagtttgaatcttttatttttttcttcc	421
Db	136081	AGGAACAAATGGGGGTCATTCAACAGTTTGAAATTCCTTTTATTTTCTTTTCC	136022
OY	422	tcaatcccttttattttttaaataagtctctttgttaatgtgtgtcaaacggaatg	481
Db	136021	TCATTCCTTTCTCATTTTAAAAAATGATCTTTTGTATGTGGGTGTTAAAAACGGAAATG	135962
OY	482	aaaactgycaccccactctcttgaacaacatcgtgaatttgaattctatgttctattatc	541
Db	135961	AAAACGTGGACCCCATCTCTTTGAAACATCTGTAATCTGAATTCATGCTCATTAATTC	135902
OY	542	attattgttgtttcatatgtatgtgcgatattttgtgtatcaagctcagtcaccttaatt	601
Db	135901	ATTATTTGTTTATTTTTCATATTGCTGCAATTTTGGATCAAGCCCTCATGCTCTTCAATAT	135842
OY	602	accctcccttctttaaataatcactgtgacaagagagtgtaaccttttcagacatctga	661
Db	135841	ACCCTCTCTTTTAAAAATTAAGCTGTGCACAGAGAGCCACCTTTTTCAGGACATTTGA	135782
OY	662	tttcaagctgtgtgtgtataataagatcgacaatgcaatgtatcataatgacttcca	721
Db	135781	TTTTCAAGCTTGTGTGTATTAATTAAGATTAACCAATCCAAATCCATTAATTAATCACTTGCA	135722
OY	722	attgcccctgatgttc-agcatgtatctacttcaactccttgcacttgcacttcaagtga	780
Db	135721	ATTGCCCTGTGATGTTTCAAGCATGTGATTACTTCACTCTCTGACTGTGACTTTCACTGGA	135662
OY	781	gatgaaattttttaaaggaactg-aactgtggaaaaaatgaacttcttcaacttgaact	839
Db	135661	GATGGAAGTGTTTTCAAGAACTGAAATGTGGAATAATGAACCTTCTTCACTTCACTGAACT	135602
OY	840	actttaaaatttgaaggtctgacccaagaagaggaataatcaagttgaattgaactg	899
Db	135601	ACTCTT-AAATTTGAGGGCTGTGCACCAAAAGGAAGGAATATTCGGGTTGAATCAAGTA	135543
OY	900	acagataaggtgagagaatgactaaactccaaaagatgtgcttcaactgaaagaagcatt	959
Db	135542	ACTGATTAAGGAGAGTAATGACTTAATCCAAAGATGAGCCTTCAATTGAGAAAGACATTT	135483
OY	960	taagaattttttaaanaactgtctcagaagaatcccaagaagatcttaatttcaattgcat	1019
Db	135482	TAAATTTTTTAAAAAATCTGTGCAGAAAGATCCAGAAAGTTCTTAATTTTCAATTACGAAT	135423

QY	1020	tataaagctatcacatgcagaaatgaatacaacagaaacacgcctttttgatattat	1079
Db	135422	TATTAAGCTATACATGCGAAGATGAAATGCAATGCAACGACACACGCTCTTTT	GATTTTAT
QY	1080	tgtactcttttggcctcgggatgatgggtctttaaataagacatgctctgtaccagctcaat	1139
Db	135363	TCGACTTTTGGCCGCGGGATGATGGGGTTTAAATGACATGTCGTGTCACCTTCATTTAA	135304
QY	1140	aataacaataatctgtccaataa	1161
Db	135303	AATAAACATATTTGTAAAAA	135282
RESULT	7		
LOCUS	AL359981/c		
DEFINITION	AL359981 167440 bp DNA	HTG	20-JAN-2001
ACCESSION	Homo sapiens chromosome 1 clone RP11-115N23, *** SEQUENCING IN		
VERSION	AL359981		
KEYWORDS	AL359981.10 GI:12331082		
SOURCE	HTG: HTGS_PHASE1; HTGS_DRAFT.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
JOURNAL	1 (bases 1 to 167440)		
COMMENT	Pavitt,R. Direct Submission Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk On Jan 22, 2001 this sequence version replaced gi:10186747.		

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

[illegible]

Query Match	68.3%;	Score 1000.2;	DB 97;	Length 1017;
Best Local Similarity	99.6%;	Pred. No. 1.8e-183;		
Matches 1013;	Conservative 0;	Mismatches 3;	Indels 1;	Gaps 1.

QY	861	gggcacaaagaagagaaatatacagttgaactcaaatatcagataaagttagaagtaatg	920
Db	841	GGACCAAAAAGAGAGAAATACAGTTGAAGTCAAGTCAAGATTAAGTGAAGTAATG	900
QY	921	aactactccaagaatgaggtttactgagaagaaagcatttaagaatltttttaaanaacttg	980
Db	901	ACTAACCTCCAAAGATGCGTTACTGCAAGAAAGGSCATTTTAAGATTTTTTAAAAATCTTG	960
QY	981	tcaagaagatcccaagaagaattctaatatlltcattagcaaatataaagctataacatgc	1037
Db	961	TCAAGAAAGATCCCAAGAAAGCTTTAATTTTCATTAGGCAAAATTAATAAGCTATTAACATGC	1017

RESULT	9
LOCUS	AC005220
DEFINITION	AC005220 163196 bp DNA PRI 01-JUL-1998 Homo sapiens chromosome 20, BAC clone 99 (LBNL H80), complete sequence.
ACCESSION	AC005220
VERSION	AC005220.1 GI:3282159
KEYWORDS	HFG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 163196)
AUTHORS	Gray,J.W., Collins,C., Kimerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Miguel,T., Miller,C., Pritchuck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H
TITLE	Sequencing of human chromosome 20
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 163196)
AUTHORS	Ricke,D.O.
TITLE	Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 163196)
AUTHORS	Gray,J.W., Collins,C., Kimerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Miguel,T., Pritchuck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H
TITLE	Direct Submission
JOURNAL	Submitted (01-JUL-1998) Human Genome Center, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, MS 74-157, Berkeley, CA 94720, U.S.A.
COMMENT	Sequence submitted by: DOE Joint Genome Institute.
FEATURES	Location/Qualifiers 1..163196 /organism="Homo sapiens" /db_xref="taxon:9606" /map="20" /clone="BAC 99" /chromosome="20" /note="LBNL H80" 4588..4709 /rpt_family="MR" 6885..6813 /rpt_family="Tiger1" complement(6838..7220) /rpt_family="LI" 7175..7314 /rpt_family="LI" 7322..7441 /rpt_family="Tiger1" complement(8695..9257) /rpt_family="MT2B2" 10121..10350 /rpt_family="Alu" 11057..11357 /rpt_family="Alu" 13431..13702 /rpt_family="Alu" 15911..15962

```
misc_feature      /rpt_family="MIR"  
16029..16484  
/note="99% identity B54864 BAC end sequence"  
repeat_region    /rpt_family="Alu"  
complement(16891..17025)  
repeat_region    /rpt_family="Alu"  
complement(17221..17682)  
repeat_region    /rpt_family="L1"  
complement(18500..18756)  
misc_feature     /rpt_family="L1"  
complement(18909..19161)  
repeat_region    /note="GRAIL 2 excellent exon, frame 1"  
19640..19943  
/rpt_family="Alu"  
complement(22025..22314)  
repeat_region    /rpt_family="Alu"  
complement(22903..23020)  
misc_feature     /note="GRAIL 2 excellent exon, frame 1"  
complement(23247..23567)  
repeat_region    /rpt_family="Alu"  
complement(24441..24514)  
misc_feature     /note="GRAIL 2 excellent exon, frame 0"  
25506..25525  
/note="(T)20"  
repeat_region    /rpt_type=tandem  
/rpt_unit="T"  
complement(25538..25679)  
repeat_region    /rpt_family="MIR"  
complement(26920..27577)  
repeat_region    /rpt_family="LTR8"  
complement(28412..28563)  
repeat_region    /rpt_family="MER20"  
28734..29019  
/rpt_family="Alu"  
repeat_region    30266..30455  
/rpt_family="THE1"  
30789..30888  
misc_feature     /note="GRAIL 2 excellent exon, frame 2"  
31841..31972  
misc_feature     /note="GRAIL 2 excellent exon, frame 0"  
32270..32290  
/note="(T)21"  
repeat_region    /rpt_type=tandem  
/rpt_unit="T"  
32702..32860  
misc_feature     /note="GRAIL 2 excellent exon, frame 1"  
33268..33474  
misc_feature     /note="GRAIL 2 excellent exon, frame 0"  
33836..33863  
/note="(GT)14"  
repeat_region    /rpt_type=tandem  
/rpt_unit="GT"  
35158..35265  
misc_feature     /note="GRAIL 2 excellent exon, frame 0"  
complement(41855..42137)  
repeat_region    /rpt_family="Alu"  
43253..43465  
/note="GRAIL 2 excellent exon, frame 1"  
complement(50487..50600)  
repeat_region    /rpt_family="L1"  
50605..50855  
repeat_region    /rpt_family="L1"  
51461..51822  
repeat_region    /rpt_family="THE1"  
complement(54343..54856)  
repeat_region    /rpt_family="LTR12"  
complement(55490..55796)  
repeat_region    /rpt_family="Alu"  
55900..56027  
repeat_region    /rpt_family="MIR"  
complement(61007..61096)  
misc_feature     /note="GRAIL 2 excellent exon, frame 1"  
join(61724..61846,71254..71319,75055..75261)  
prim_transcript  /standard_name="af74a01"
```

```
repeat_region    /note="100% identity dbEST:AA644590"  
complement(62247..62746)  
repeat_region    /rpt_family="L1"  
complement(65258..65541)  
repeat_region    /rpt_family="Alu"  
65587..65882  
repeat_region    /rpt_family="Alu"  
68256..68282  
/note="(T)27"  
repeat_region    /rpt_type=tandem  
/rpt_unit="T"  
69077..69495  
repeat_region    /rpt_family="Alu"  
69537..69618  
repeat_region    /rpt_family="Alu"  
complement(70542..70721)  
repeat_region    /rpt_family="MERS"  
72072..72255  
repeat_region    /rpt_family="MIR"  
complement(72260..72340)  
repeat_region    /rpt_family="MLT1"  
74417..74456  
repeat_region    /note="(CT)20"  
/rpt_type=tandem  
/rpt_unit="CT"  
75057..75320  
misc_feature     /note="GRAIL 2 excellent exon, frame 2"  
join(75199..75323,84601..84746)  
prim_transcript  /standard_name="zm96b07"  
/note="96% identity dbEST:AA079436"  
complement(77059..77202)  
repeat_region    /rpt_family="MERS"  
complement(77269..77551)  
repeat_region    /rpt_family="Alu"  
join(78685..79043,88203..88578)  
prim_transcript  /standard_name="ym20b03"  
/note="97% identity dbEST:HL15040"  
79526..79590  
prim_transcript  /standard_name="nc17d10"  
/note="81% identity dbEST:AA259187"  
complement(81547..81627)  
repeat_region    /rpt_family="MIR"  
82700..82865  
repeat_region    /rpt_family="Alu"  
83191..83347  
repeat_region    /rpt_family="L1"  
complement(83468..83594)  
repeat_region    /rpt_family="MIR"  
84144..84436  
repeat_region    /rpt_family="Alu"  
complement(84402..84774)  
prim_transcript  /standard_name="nl14b12"  
/note="98% identity dbEST:AA554549"  
complement(85065..85162)  
prim_transcript  /standard_name="yw21h10"  
/note="90% identity dbEST:H88501"  
join(85165..85589,86249..86579,87354..87476)  
prim_transcript  /standard_name="yu90d02"  
/note="97% identity dbEST:RA3020"  
complement(86451..86577)  
misc_feature     /note="GRAIL 2 excellent exon, frame 2"  
complement(87953..88445)  
prim_transcript  /standard_name="zu90c12"  
/note="99% identity dbEST:AA626177"  
complement(89695..89866)  
repeat_region    /note="GRAIL 2 excellent exon, frame 1"  
89959..90146  
repeat_region    /rpt_family="MERS"  
92028..92411  
Query Match      55.5%; Score 813; DB 85; Length 163196;  
Best Local Similarity 88.0%; Pred. No. 2,8e-147;  
Matches 1032; Conservative 0; Mismatches 105; Indels 36; Gaps 12;
```


QY 248 gacagggtgttccaatgaatcaccacaggttctcttcttgagggtcagaagaatgtcgtgata 307
 |||||
 Db 241 GACAGGGGTTCACAAATGAATTCACCTCAGGTTTCTTTGAGGGTCAGAGAAATGCTGATA 300
 QY 308 atcaactcacaagaacatcggaatgagagaagaagtgtgatgaaqgttatacagaac 367
 |||||
 Db 301 ATCATATCCAAAGAACTGGAGATGAGAGAAAGATGTCATTGAAGTTATCAGGAAAC 360
 QY 368 aaagaggaggtcattcaacaggttagatattcttatttttttttttttttttttttttttt 427
 |||||
 Db 361 AACGGGGGGGATTCACACAGTTAGATATTCCTTTATTTTCTTTTCTTTCCCTCAATC 420
 QY 428 ctttttattttttaaataagttcttcttcttcttcttcttcttcttcttcttcttcttctt 487
 |||||
 Db 421 CTTTATTTTATTTTAAATAAGTCTTTTGTATGNGGTTCACAAACGGAATGMAAACT 480
 QY 488 ggaaccacatctcttgaacaacatctgtgaattgtatctctctctctctctctctctct 547
 |||||
 Db 481 GGCACCCCATCTCTTTGAAACATCTGTAAATTTGAATTCATGTCCTCATATTCATTAT 540
 QY 548 gttgttttcaattgtcgtatcttcttcttcttcttcttcttcttcttcttcttcttctt 607
 |||||
 Db 541 GTTGTATTTTCAATGTCGTGATTTTGGTATCAGCAACCTCAGTCCCTTCATATTAACCTC 600
 QY 608 tcccttttaaaataatcagctgtgacagagaagtcaccccttcttcttcttcttcttcttct 667
 |||||
 Db 601 TCCCTTTTAAATAATTCAGTGTGACAGAGAGTCCCTTTTTCAGAGATTCATTTTCA 660
 QY 668 gacttggtgtgataataatgacagacacacacacacacacacacacacacacacacacac 727
 |||||
 Db 661 GCGTTGTGTGTAATTAATGATGACCAATGCAATGTCATTAATGATTCATTAATGATGAC 720
 QY 728 cctgatgttc-agcagatgtatcttcttcttcttcttcttcttcttcttcttcttcttct 785
 |||||
 Db 721 CCGTATGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 QY 786 aagtttttc--agagaactgaactgtgaaaaatga 819
 |||||
 Db 781 AATTTTTCAGAAAAAAGTGAACGTGTGAAAAATGA 816
 RESULT 11
 AF033353 1187 bp mRNA ROD 06-FEB-1998
 LOCUS AF033353 ubiquitin-homology domain protein (Ub1) mRNA,
 DEFINITION complete cds.
 ACCESSION AF033353
 VERSION AF033353.1 GI:2645736
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1187)
 AUTHORS Howe, K., Williamson, J., Boddie, N., Sheer, D., Freemont, P. and
 Solomon, E.
 TITLE The ubiquitin-homology gene P1C1: characterization of mouse (P1C1)
 and human (UB1) genes and pseudogenes
 JOURNAL Genomics 47 (1), 92-100 (1998)
 MEDLINE 98126440
 REFERENCE 2 (bases 1 to 1187)
 AUTHORS Howe, K., Freemont, P.S. and Solomon, E.
 TITLE Direct Submission
 JOURNAL Submitted (07-NOV-1997) Medical & Molecular Genetics, Guy's
 Hospital, 8th floor, Guy's Tower, London SE19RT, UK
 FEATURES
 source
 1. 1187
 /organism="Mus musculus"
 /strain="ICR"
 /DB_xref="taxon:10090"
 /chromosome="1"
 /map="C2-3"
 1. 1187
 gene

CDS
 /gene="ub1"
 /note="P1C1"
 119..424
 /gene="ub1"
 /codon_start=1
 /product="ubiquitin-homology domain protein"
 /protein_id="AAC39959.1"
 /db_xref="GI:2645737"
 /translation="MSDPAKSTEDLDKKEGEYIKIKYIGDSSSEIHFVKMTTHL
 KRLESYCRQGVPMNSLRFLFEGRIADNHTPKELMEDEEDYIEVYDQIGHSIV"
 BASE COUNT 350 a 216 c 267 g 354 t
 ORIGIN
 Query Match 52.3%; Score 765.8; DB 94; Length 1187;
 Best Local Similarity 84.4%; Pred. No. 3.2e-138;
 Matches 988; Conservative 0; Mismatches 152; Indels 31; Gaps 10;
 2 gggaagattgttaaaccccgagcgaggtctgtctaccgcgagcgctgtgtcgga 61
 |||||
 Db 34 GGGACGGATTGTAAACCTCAGATGAGTTCGCTGCCGAGAGCGCAGTGTGCGGA 93
 QY 62 gaccgccgggtgaagccacgcatcatcgtcgtcgaagaaggaacacacacacacacacacac 121
 |||||
 Db 94 GACCCCGGGGGAATCCA-CGTACCATGTCGTGACAGAGCAAAACCTTCAACTGAG 152
 QY 122 acttgggggataagaacgaagtgatataataactcaaatgattgtgacagatagca 181
 |||||
 Db 153 ACTTAGCGATTAAGAAAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 212
 QY 182 gtgagttcacttaaaagtgaataacacacacacacacacacacacacacacacacacacac 241
 |||||
 Db 213 GTGAGATATATTAAAGTGAATGACAAACATCTCAAGAAACCTCAAGATCATATCT 272
 QY 242 gtcaagaacaggtgttccaaatgaatcactcaaggttcttcttcttcttcttcttcttct 301
 |||||
 Db 273 GTCAAGACAGAGGATTCATTAATGATTCACACAGTTCCTTTGAAGGTGAGGAATTTG 332
 QY 302 ctgataatcatatcccaaaagaacttggaatgagagaagaagtgtatgaagtatc 361
 |||||
 Db 333 CTGATTAATCATCTCCGAAAGACTGGGAATGAGAGAAAGATGATGAAGTTATC 392
 QY 362 aggaacaaacgggggtcatcacaacagtttagata----tcttttatttttttctt 416
 |||||
 Db 393 AGGAACAAACGGGGGTCTACTCGACGGTTAGATTAATCTTTTATTTTATTTTCT 452
 QY 417 ttcctcaatccttttattttaaataatgtcttcttcttcttcttcttcttcttcttcttct 475
 |||||
 Db 453 TTCCCTCAATCTTTTATTTTAAATAATGTTCTTTGTAATGTGTGTCATAAATGA 512
 QY 476 gaattgaacactgaccccatcctcttgaacacatctgtaattgaattgaattgaattga 535
 |||||
 Db 513 AATTTGAATCTGACATCTCTTTGAA-----CATATGAATCTGATGTTG 563
 QY 536 ttatcatatgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 595
 |||||
 Db 564 AATTCATTAATGTTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 623
 QY 596 catatcactccttcttctttaa---aaattcgtgtgacagagaggtcacttcttcaag 652
 |||||
 Db 624 AATTAATGCTTTTCCCTTTTAAAGATTTTCAATGTGTGACAGAGAGGCCACCTTTTCAG 683
 QY 653 gacattgacatttcagagctgtgtgataaataagatcgacaatgcaagtgttcaataat 712
 |||||
 Db 684 GACTGTGCAATTTTCAAGTTTGTGATGATAAA-AAGATGACCAATGAGAGTTCCTCATG 742
 QY 713 gacttccaattgcccctgagatgttaagatgtgatacttaactccttcttcttcttcttct 772
 |||||
 Db 743 ACCTTCAATGTGCTGAAAGTTCACAGCATGTGTTGCTTCACTCTGACATGATATTT 802
 QY 773 cagtgagagatgagagtttttcagagaactgaactgtgaaataatgaacttcttcttact 832
 |||||
 Db 803 CAGTGGAGATGGAATTTTTCAGAGAACTGAACGTGTG-AAATGACCTTTCCACACT 861

QY 833 tgaagctactttaaattgagggtcgtgacaaagaagaataatcagttgaagt 892
 Db 862 TGAAGCTACTTTTAAATCTCGGGCTGACCAAAAGAACATCAT--GTTTGTAGT 919
 QY 893 caagatgacagataagtgagagatgaatgaactccaagaatgctcactgaaga 952
 Db 920 CAAGATGACAGATACAGAGACAGACAGTAACTCCAAAGTGGCTTCATGAGAGAA 979
 QY 953 ggcatttaagatttttaaaatctgtctgagaagatccagaagtttcaattcat 1012
 Db 960 AGAAGTCTCTTGAACAAGACAGATGCTGAGAGATCCAGAAAGTTCTAATGTTTCA 1039
 QY 1013 taaga--attaataaagctatacatcagaatgaatacacagaacactgcctttt 1069
 Db 1040 CAGCAGTTATTAAATTAAGTTACTATACAGAGTGTACGACACAGA-----CACTGCTTT 1094
 QY 1070 agatttatttgaacttttgctgagatagggtttaatgacattgtctgacca 1129
 Db 1095 TGATTCTGTTTGTACTTTTGGCTGGACATGAGTGTTCAGGAGACATCOTCTGTACA 1154
 QY 1130 gcttcaataaataaacaatatgttcaaaa 1160
 Db 1155 GCTTCATTAAATTAACAATATTGTATAAAA 1185

RESULT 12
 AC084066/c DNA HTG 12-OCT-2000
 LOCUS AC084066.1 GI:10799415
 DEFINITION Mus musculus clone RP23-321D1, *** SEQUENCING IN PROGRESS ***
 ACCESSION AC084066
 VERSION AC084066.1
 KEYWORDS HTG; HTGS; PHASE1.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 235411)
 AUTHORS DOE Joint Genome Institute.
 TITLE DOE Joint Genome Institute.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 235411)
 AUTHORS DOE Joint Genome Institute.
 TITLE DOE Joint Genome Institute.
 JOURNAL Direct Submission
 COMMENT Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

 Project Information
 Center Project Name: 2351294
 Center clone name: RP23-321D1

Summary Statistics
 Consensus quality: 214207 bases at least Q40
 Consensus quality: 223053 bases at least Q30
 Consensus quality: 225208 bases at least Q20
 Estimated insert size: 200000; pulse field gel estimation
 Estimated insert size: 232611; sum-of-contigs estimation
 Quality coverage: 11.56 in Q20 bases; pulse field gel estimation
 Quality coverage: 9.94 in Q20 bases; sum-of-contigs estimation
 NOTE: This is a 'working draft' sequence. It currently
 consists of 29 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1210: contig of 1210 bp in length

FEATURES
 SOURCE
 1.235411
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-321D1"
 /clone_lib="RP23 mouse BAC library 23"
 BASE COUNT 62132 a 54445 c 54225 g 61806 t 2803 others
 ORIGIN

Query Match 50.8%; Score 743.6; DB 76; Length 235411;
 Best Local Similarity 83.6%; Pred. No. 7e-134;
 Matches 971; Conservative 0; Mismatches 159; Indels 32; Gaps 10;

QY 12 tgaataaccgagagaggtctgcttaaccgagcgctgctgtgagagaccccgag 71
 Db 113180 TGAATAACCTCAGAGTGTCTGCTGCCGAGGCCGACAGTGTGTCAGAGACCCCGGG 113121

Thu Aug 16 08:48:49 2001

us-09-484-964-1.rge

Page 20

QY 1129 agcttcattaaataaacacatattgtcaca 1160
|||||
Db 121256 AGCTTCATTAAATAAACACATATTGTGTA 121287

Search completed: August 15, 2001, 16:06:02
Job time: 3108 sec

	Search time	Cell updates/sec
search, using sw model	131.67 seconds	
without alignments)		

...catgccagcc 1465

Gapor 313950809 Resu- 1460202

of hits satisfying

length: 0

Series	Match	0%	100%
Minimum	Match	0%	100%

Maxim first
Listing

```

1: /SID1/gcgdata/geneseq/geneseq.NA1984.DAT:
2: /SID1/gcgdata/geneseq/geneseq.NA1984.DAT:
3: /SID1/gcgdata/geneseq/geneseq.NA1984.DAT:

```

SUMMARY

Description

[illegible]

ALIGNMENTS

Aspergillus ...

XX
DR
WPI; 1998-2000

DE Human secreted protein gene 47 clone HOGAV75.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 XX MO9839448-A2.
 XX
 PD 11-SEP-1998.
 XX
 PF 06-MAR-1998; 98WO-US04493.
 XX
 XX 02-OCT-1997; 97US-0061060.
 PR 07-MAR-1997; 97US-0038621.
 PR 07-MAR-1997; 97US-0040161.
 PR 07-MAR-1997; 97US-0040162.
 PR 07-MAR-1997; 97US-0040163.
 PR 07-MAR-1997; 97US-0040333.
 PR 07-MAR-1997; 97US-0040334.
 PR 07-MAR-1997; 97US-0040336.
 PR 07-MAR-1997; 97US-0040626.
 PR 11-APR-1997; 97US-0043311.
 PR 11-APR-1997; 97US-0043312.
 PR 11-APR-1997; 97US-0043313.
 PR 11-APR-1997; 97US-0043314.
 PR 11-APR-1997; 97US-0043568.
 PR 11-APR-1997; 97US-0043569.
 PR 11-APR-1997; 97US-0043576.
 PR 11-APR-1997; 97US-0043580.
 PR 11-APR-1997; 97US-0043580.
 PR 11-APR-1997; 97US-0043670.
 PR 11-APR-1997; 97US-0043671.
 PR 11-APR-1997; 97US-0043672.
 PR 11-APR-1997; 97US-0043674.
 PR 23-MAY-1997; 97US-0047492.
 PR 23-MAY-1997; 97US-0047500.
 PR 23-MAY-1997; 97US-0047501.
 PR 23-MAY-1997; 97US-0047502.
 PR 23-MAY-1997; 97US-0047503.
 PR 23-MAY-1997; 97US-0047581.
 PR 23-MAY-1997; 97US-0047582.
 PR 23-MAY-1997; 97US-0047583.
 PR 23-MAY-1997; 97US-0047584.
 PR 23-MAY-1997; 97US-0047585.
 PR 23-MAY-1997; 97US-0047586.
 PR 23-MAY-1997; 97US-0047587.
 PR 23-MAY-1997; 97US-0047588.
 PR 23-MAY-1997; 97US-0047589.
 PR 23-MAY-1997; 97US-0047590.
 PR 23-MAY-1997; 97US-0047592.
 PR 23-MAY-1997; 97US-0047593.
 PR 23-MAY-1997; 97US-0047594.
 PR 23-MAY-1997; 97US-0047595.
 PR 23-MAY-1997; 97US-0047596.
 PR 23-MAY-1997; 97US-0047597.
 PR 23-MAY-1997; 97US-0047598.
 PR 23-MAY-1997; 97US-0047599.
 PR 23-MAY-1997; 97US-0047600.
 PR 23-MAY-1997; 97US-0047612.
 PR 23-MAY-1997; 97US-0047613.
 PR 23-MAY-1997; 97US-0047614.
 PR 23-MAY-1997; 97US-0047615.
 PR 23-MAY-1997; 97US-0047617.
 PR 23-MAY-1997; 97US-0047618.

PR 23-MAY-1997; 97US-0047632.
 PR 23-MAY-1997; 97US-0047633.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048974.
 PR 08-JUL-1997; 97US-0049610.
 PR 16-JUL-1997; 97US-0051926.
 PR 18-AUG-1997; 97US-0052874.
 PR 22-AUG-1997; 97US-0055724.
 PR 22-AUG-1997; 97US-0056630.
 PR 22-AUG-1997; 97US-0056631.
 PR 22-AUG-1997; 97US-0056632.
 PR 22-AUG-1997; 97US-0056636.
 PR 22-AUG-1997; 97US-0056637.
 PR 22-AUG-1997; 97US-0056662.
 PR 22-AUG-1997; 97US-0056664.
 PR 22-AUG-1997; 97US-0056845.
 PR 22-AUG-1997; 97US-0056862.
 PR 22-AUG-1997; 97US-0056864.
 PR 22-AUG-1997; 97US-0056872.
 PR 22-AUG-1997; 97US-0056874.
 PR 22-AUG-1997; 97US-0056875.
 PR 22-AUG-1997; 97US-0056876.
 PR 22-AUG-1997; 97US-0056877.
 PR 22-AUG-1997; 97US-0056878.
 PR 22-AUG-1997; 97US-0056879.
 PR 22-AUG-1997; 97US-0056880.
 PR 22-AUG-1997; 97US-0056881.
 PR 22-AUG-1997; 97US-0056882.
 PR 22-AUG-1997; 97US-0056884.
 PR 22-AUG-1997; 97US-0056886.
 PR 22-AUG-1997; 97US-0056887.
 PR 22-AUG-1997; 97US-0056888.
 PR 22-AUG-1997; 97US-0056889.
 PR 22-AUG-1997; 97US-0056892.
 PR 22-AUG-1997; 97US-0056893.
 PR 22-AUG-1997; 97US-0056903.
 PR 22-AUG-1997; 97US-0056908.
 PR 22-AUG-1997; 97US-0056909.
 PR 22-AUG-1997; 97US-0056910.
 PR 22-AUG-1997; 97US-0056911.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057669.
 PR 05-SEP-1997; 97US-0057761.
 PR 12-SEP-1997; 97US-0058785.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
 PI Peng P, Ferlie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 DR WPI; 1998-506364/43.
 DR P-PsDB; AAW74777.
 XX
 PT New isolated human genes and the secreted polypeptide(s) they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1: Page 279-280; 721pp; English.
 XX
 CC This sequence represents a nucleic acid molecule designated Gene 47 from
 CC the human cDNA clone HOGAV75 (deposited as clone ATCC 97899 and ATCC
 CC 209045) which encodes a secreted human protein. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
 CC the fused protein as compared to the human protein only.
 CC The invention relates to 186 novel genes and their fragments (nucleic
 CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W5026)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological conditions
 CC can be diagnosed by determining the amount of the new polypeptides in a

CC sample or by determining the presence of mutations in the new
CC polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAV59511 for described uses).

sequence 1220 BP; 382 A; 207 C; 258 G; 372 T; 1 other

Query Match	77.28;	Score 1131.2;	DB 19;	Length 1220;
Best Local Similarity	99.68;	Pred. No. 9,'e-24;		
Matches 115;	Conservative	0;	Mismatches 3;	Indels 2;
				Gaps 2.

[illegible]

Accession	Sequence	Length
Db	942 cagataagggaagtaagtaacttaactccaagaatggtctcactgaaagaagacattt	1001
Qy	961 aagatttttaaaaactctgtcagaagatccccaagaagttcatttcattgacat	1020
Db	1002 aagatttttaaaaactctgtcagaagatccccaagaagttcatttcattgacat	1061
Qy	1021 aataaagctataatgcagaaatgtaatacacagaacactgtctcttttaatttatt	1080
Db	1062 aataaagctataatgcagaaatgtaatacacagaacactgtctctttt-gatttaatt	1120
Qy	1081 gtacttttgccctgagatagtggttttaaatgagacattgtctgtaacagcttcaataa	1140
Db	1121 gtacttttgccctgagatagtggttttaaatgagacattgtctgtaacagcttcaataa	1180
Qy	1141 ataacaataattgtcaaaa 1160	
Db	1181 ataacaataattgtcaaaa 1200	

RESULT 4

AAV59122
ID AAV59722 standard; DNA; 1196 BP

AAV59722 AC

DT 19-JAN-1999 (first entry)

XX
DE Human secreted protein gene 47 clone HOGAV75.

XX Human; secreted protein; fusion protein; gene therapy; protein leukemia;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental disease; allergy; foetal deficiency; blood; allergy; renal; ds;
KW immune system; lymphoma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischemic shock; Alzheimer's disease; osteoclasts; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; Thymus
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW osteoporosis; metabolism; regulation; malabsorption; gastritis; neoplasm.

aa Homo sapiens

PN W09839448-A2

PD 11-SEP-1998

PF 06-MAR-1998; 98WO-US04493

PR	02-OCT-1997	97US-0061060
PR	07-MAR-1997	97US-0038621
PR	07-MAR-1997	97US-0040161
PR	07-MAR-1997	97US-0040162
PR	07-MAR-1997	97US-0040163
PR	07-MAR-1997	97US-0040333
PR	07-MAR-1997	97US-0040334
PR	07-MAR-1997	97US-0040336
PR	07-MAR-1997	97US-0040626
PR	07-MAR-1997	97US-0043311
PR	11-APR-1997	97US-0043312
PR	11-APR-1997	97US-0043313
PR	11-APR-1997	97US-0043314
PR	11-APR-1997	97US-0043568
PR	11-APR-1997	97US-0043569
PR	11-APR-1997	97US-0043576
PR	11-APR-1997	97US-0043578
PR	11-APR-1997	97US-0043580
PR	11-APR-1997	97US-0043589
PR	11-APR-1997	97US-0043670
PR	11-APR-1997	97US-0043671
PR	11-APR-1997	97US-0043672
PR	11-APR-1997	97US-0043674
PR	23-MAY-1997	97US-0047592
PR	23-MAY-1997	97US-0047590
PR	23-MAY-1997	97US-0047501

XX	PA	XX	(HUMA-)	HUMAN GENOME SCI INC.
PI	Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,			
PI	Peng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,			
PI	Kyaw H, Latleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,			
PI	Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;			
DR	WPI: 1998-506364/43.			
DR	P-PDB: AAM74938.			
PT				
PT	New isolated human genes and the secreted polypeptide(s) they encode			
PT	- useful for diagnosis and treatment of e.g. cancers, neurological			
PT	disorders, immune diseases, inflammation or blood disorders			
XX	Claim 1: Page 449: 721pp: English.			
CC				
CC	This sequence represents a nucleic acid molecule designated Gene 47 from			
CC	the human cDNA clone HOGAV75 (deposited as clone ATCC 97897 and ATCC			
CC	20904/3) which encodes a secreted human protein. The gene can be used to			
CC	generate fusion proteins by linking to the gene to a human			
CC	immunoglobulin Fc portion (e.g. AAV5502) for increasing the stability of			
CC	the fused protein as compared to the human protein only.			
CC	The invention relates to 186 novel genes and their fragments (nucleic			
CC	acid sequences: AAV59511-595812; amino acid sequences AAM74731-W75026)			
CC	which are useful for preventing, treating or ameliorating medical			
CC	conditions e.g. by protein or gene therapy. Also, pathological			
CC	polypeptides in a sample or by determining the amount of the new			
CC	the new polynucleotides. Specific uses are described for each of the 186			
CC	(see AAV5511 for described uses).			
XX				
XX	Sequence 1196 BP: 360 A; 212 C; 259 G; 364 T; 1 other;			
QY				
QY	Query Match	76.6%;	Score 1122.8;	DB 19; Length 1196;
QY	Best Local Similarity	99.7%;	Prod. No. 7.3e-242;	
QY	Matches 1146; Conservative	0;	Mismatches 2;	Indels 2; Gaps 2
QY	2	gggagagattgtgtaaacccgcgagcggtctgtgtaaccgcgagcgccgtcgtgcgga	61	
QY	48	gagagagattgttaaacccgcgagcggtctgtgtaaccgcgagcgccgtcgtgcgga	107	
QY	62	gaccgccgggtgtagccacgcgtcatcatgtctgcaccagaggcgaaaccttaactgag	121	
QY	108	gaccccggtgtgaagcaccgcgtcatcatgtctgcaccagaggcgaaaccttaactgag	167	
QY	122	acttgggggttaagcaagcaggtgtaatatatttaactcaaatgcatcttggacagataag	181	
QY	168	acttgggggttaagcaagcaggtgtaatatatttaactcaaatgcatcttggacagataag	227	
QY	182	gtgagattcaactcaaatgtaaatgacacacatctcaagaactcaagaatatatt	241	
QY	228	gtgagattcaactcaaatgtaaatgacacacatctcaagaactcaagaatatatt	287	
QY	242	gtcaagaagcaggtgttccaatgaaatcaactagttctcttggagggtcagagaattg	301	
QY	288	gtcaagaagcaggtgttccaatgaaatcaactagttctcttggagggtcagagaattg	347	
QY	302	ctgataatcatctccaaagaactgggaatggaggaagaatgtgataagtttacc	361	
QY	348	ctgataatcatctccaaagaactgggaatggaggaagaatgtgataagtttacc	407	
QY	362	aggacaacacgggggtcatctcaacagtttgatattcttttattttttttcc	421	
QY	408	aggacaacacgggggtcatctcaacagtttgatattcttttattttttttcc	467	
QY	422	tcaatcctttttatttttaaaaatgattccttggtaatgggtgtcacaacaggaattg	481	
QY	468	tcaatcctttttatttttaaaaatgattccttggtaatgggtgtcacaacaggaattg	527	

```

DB 528 aaactggcaccacccctccttgaacacccggaattcgaatcagtcacattc 587
QY 542 attatgtttgtttcattgctgctgattttgttgatcaacccctcattcatt 601
DB 588 attatgtttgtttcattgctgctgattttgttgatcaacccctcattcatt 647
QY 602 accctcctcttttaaaaaattacgtgtgcaagagaggtcaccttttcaggacatgca 661
DB 648 accctcctcttttaaaaaattacgtgtgcaagagaggtcaccttttcaggacatgca 707
QY 662 tttttaggtgtgtgtttaaataagatccgacaaatgcaaggttcataaattcattc 721
DB 708 ttcttaggtgtgtgtttaaataagatccgacaaatgcaaggttcataaattcattc 767
QY 722 attgcccctgagatc-aggatgtattactcactcctgagctgtgacttcagtgga 780
DB 768 attgcccctgagatc-aggatgtattactcactcctgagctgtgacttcagtgga 827
QY 781 gatggaagttttcagagagactggaactgtggaataatgacatttccttaacttgaagcta 840
DB 828 gatggaagttttcagagagactggaactgtggaataatgacatttccttaacttgaagcta 887
QY 841 ctttaaaatttgagaggtgtggaacaaagaaggaataatgaggttgaagtaagatgga 900
DB 888 ctttaaaatttgagaggtgtggaacaaagaaggaataatgaggttgaagtaagatgga 947
QY 901 cagataagtgagagatgacttaactccaagaatggtctcactgaagaagaagatttt 960
DB 948 cagataagtgagagatgacttaactccaagaatggtctcactgaagaagaagatttt 1007
QY 961 aagattttttaaataatctgtcagaagaatcccaagaagaatgttaatttcattgaacatt 1020
DB 1008 aagattttttaaataatctgtcagaagaatcccaagaagaatgttaatttcattgaacatt 1067
QY 1021 aataaagcctacatcagcaaaatgaatacaacagaacacatgtcttttaagtattatt 1080
DB 1068 aataaagcctacatcagcaaaatgaatacaacagaacacatgtcttttaagtattatt 1126
QY 1081 gtactttttgagcctgagatggttttaaatgagacatgtctgtacacagcttcattaa 1140
DB 1127 gtactttttgagcctgagatggttttaaatgagacatgtctgtacacagcttcattaa 1186
QY 1141 ataacaata 1150
DB 1187 ataacaata 1196

```

RESULT 5

AAA16215 standard; DNA; 616 BP.

AC AAA16215;

DT 14-JUN-2000 (first entry)

DE Human colon cancer differentially expressed nucleotide sequence #220.

KM Colon cancer; detect; differential expression; human; treatment;
 detect mutation; non-invasive diagnostic method; ds.

OS Homo sapiens.

PN WO200012702-A2.

PD 09-MAR-2000.

PF 30-AUG-1999; 99MO-US19424.

PR 31-AUG-1998; 98US-0098639.

PR 27-JAN-1999; 99US-0117393.

PA (FARB) BAYER CORP.

```

XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
PI Catino TJ, D'Wivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
PI Schlegel R;
XX WPI: 2000-256641/22.
DR
XX
PT Novel nucleic acids and proteins for identifying therapeutic agents
PT useful for treating and diagnosing cancer, especially colon cancer
PS
XX Claim 16; Page 211-212; 345pp; English.

```

This sequence represents a human nucleotide sequence which is differentially expressed in colon cancer cells compared to the expression levels in normal cells. The nucleotide sequence can be used as a source of primers and probes. The nucleotide sequence is useful for determining the phenotype of a cell by detecting the differential expression of the sequence relative to a normal cell. The probes derived from the sequence can also be used to determine the phenotype of cells in a sample. Probes and antibodies which hybridise to the nucleotide sequence can also be used to determine the phenotype of a cell. The primers are useful for detecting a mutation in a test nucleotide sequence and also for detecting cancer, preferably colon cancer. Antibodies against the protein encoded by the nucleotide sequence can also be used in a method to detect colon cancer. The diagnostic method is non-invasive and accurate for diagnosing colon cancer at an early stage.

Sequence 616 BP; 181 A; 109 C; 149 G; 170 T; 7 other;

Query Match 32.0%; Score 469.2; DB 21; Length 616;

Best Local Similarity 95.2%; Pred. No. 7; 1e-96;

Matches 555; Conservative 0; Mismatches 20; Indels 8; Gaps 7;

```

QY 2 ggaagagattgttaaaccccgagagaggtctgtcttaaccgagccgctgtgtcgga 61
DB 41 gagaagagattgttaaaccccgagagaggtctgtcttaaccgagccgctgtgtcgga 100
QY 62 gaaccgaggtgaagacacgctcatatgtctgacacagagagcaaaccttcaactgag 121
DB 101 gaaccgaggtgaagacacgctcatatgtctgacacagagagcaaaccttcaactgag 160
QY 122 actggggagataagaagaaggtgataatattaaactaaagcttggacagataga 181
DB 161 actggggagataagaagaaggtgataatattaaactaaagcttggacagataga 220
QY 182 gtgagatcacttcaaatgaaatgacaacacatcctcaagaactcaagaatcact 241
DB 221 gtgagatcacttcaaatgaaatgacaacacatcctcaagaactcaagaatcact 280
QY 242 gtcaagaagcaggtgttccaagaatcactcaggtctctttgagggtcaggaattg 301
DB 281 gtcaagaagcaggtgttccaagaatcactcaggtctctttgagggtcaggaattg 340
QY 302 ctgataatcactccaaagaagcaggtggaatgaggaagaatgtgtg-aaagttaa 360
DB 341 ctgataatcactccaaagaagcaggtggaatgaggaagaatgtgtg-aaagttaa 400
QY 361 cagaacaacagggggtcattcaacagtttagatattcctttttttttttttttcc 420
DB 401 cagaacaacagggggtcattcaacagtttagatattcctttttttttttttttttc 459
QY 421 ctcaatccttttttttttttttttttttttttttttttttttttttttttttttt 480
DB 460 ctcaatccttttttttttttttttttttttttttttttttttttttttttttttt 518
QY 481 gaaactggcaccacatcctcttgaacaactgtgaattgaattcagtgctcattat 540
DB 519 gaaactggcaccacatc-cttgaacaactgtgaattgaattcagtgctcattat- 575
QY 541 cactatgtttgtttcattcgtgctgattttgtgtgacaaagc 583
DB 576 cactatgtttgttt--cattgacgaattttgaggatcaanc 616

```



```

Db      1 atttgtaaaccccgagagaggtctctgttaaccgagcgctgtgaggaacccc 60
Oy      69 gggfagaagccacgcgtcatcaltgcttgaccagagagcaaaccttaactgaagactgg 128
Db      61 gggfagaagccacgcgtcatcaltgcttgaccagagagcaaaccttaactgaagactgg 120
Oy      129 ggataagaagcaggtgataatataataaactcaagtcatttgacagagataagcagtggat 188
Db      121 ggataagaagcaggtgataatataataaactcaagtcatttgacagagataagcagtggat 180
Oy      189 tcaactcaaatgaaatgagacaacacacatcaagaactcaagaatataactcttcaag 248
Db      181 tcaactcaaatgaaatgagacaacacacatcaagaactcaagaatataactcttcaag 240
Oy      249 acaagggtgtccaatgaattcactcaggtttctcttgaagggtcagagaattgctgataa 308
Db      241 acaagggtgtccaatgaattcactcaggtttctcttgaagggtcagagaattgctgataa 300
Oy      309 tcaactccaaaagaagactgggaatgagagaagatgtgattgaattatcaggaaaca 368
Db      301 tcaactccaaaagaagactgggaatgagagaagatgtgattgaattatcaggaaaca 360
Oy      369 aacgggggggtca 380
Db      361 aacgggggggtca 372

```

RESULT 8
AAT03735
ID AAT03735 standard; cDNA: 425 BP.

AC AAT03735;
XX

DT 26-MAR-1996 (first entry)
XX

DE TNF-R p55IC/Fas-IC-binding protein DD11 cDNA clone.
XX

KM Tumour necrosis factor receptor; TNF-R; p55IC; Fas-IC;
KW Intracellular domain binding protein; ss.
XX

OS Homo sapiens.
XX

PN W09531544-A1.
XX

PD 23-NOV-1995.
XX

PF 11-MAY-1995; 95WO-US05854.
XX

PR 02-OCT-1994; 94IL-0111125.
XX

PR 11-MAY-1994; 94IL-0109632.
XX

PA (WEIN/) WEINMURZEL H.
XX

PA (YEDA) YEDA RES & DEV CO LTD.
XX

PI Boldin M, Mett I, Varfolomeev E, Wallach D;
XX

DR WPI; 1996-010930/01.
XX

PT TNF-NGF receptor superfamily intracellular domain-binding proteins -
XX

PT useful for modulating receptor function, e.g. for treating tumours
XX

PT or HIV-infected cells
XX

PS Claim 18; Fig 12; 96pp; English.
XX

XX cDNA clone DD11 (AAT03735) was isolated by screening a human HeLa
XX library using the human tumour necrosis factor receptor (TNF-R)
XX p55 death domain (p55D) as 'bait'. The full length of the DD11
XX transcript is approx 1.2 kb. The protein encoded by cDNA DD11
XX interacts strongly with p55D (amino acids 326-414) and also with
XX human and mouse FAS intracellular domain. Such proteins may
XX be used modulate TNF-R function.

SQ Sequence 425 BP; 135 A; 80 C; 115 G; 95 T; 0 other;

Query Match 23.6%; Score 345.2; DB 17; Length 425;
Best Local Similarity 95.7%; Pred. No. 3.3e-68;
Matches 377; Conservative 0; Mismatches 13; Indels 4; Gaps 2;

```

Oy      2 ggggaagatttgaataacccggagcgaggttctgttaccggcgccgtgtgaggga 61
Db      32 gagaagatttgaataacccggagcgaggttctgttaccggcgccgtgtgaggga 91
Oy      62 gaccccggttgaagccacgcatcatgcttgaccagagagcaaacctt-caactgg 120
Db      92 gaccccggttgaagccacgcatcatgcttgaccagagagcaaaccttcaactgg 151
Oy      121 gacttgggggataaagaacaggtgaatatataataaactcaagtcatttgacagataagc 180
Db      152 gacttgggggataaagaagaggtgaatatataataaactcaagtcatttgacagataagc 211
Oy      181 agtgaagttcacttcaagtgaaatgacacacatctcaagaactcaagaatcatatc 240
Db      212 agtgaagttcacttcaagtgaaatgacacacatctcaagaactcaagaatcatatc 271
Oy      241 tgtcaagaacaggggtgtccaatgaattcactcaggtttctcttgaagggtcagagaatt 300
Db      272 tgtcaagaacaggggtgtgtccaatgaattcactcaggtttctcttgaagggtcagagaatt 331
Oy      301 gctgataatcacttccaaaagaactggaatgagagaaga---tgtgattgaatt 357
Db      332 gctgataatcacttccaaaagaactggaatgagagaagaatgattgttgaatt 391
Oy      358 tatcagaacaacaggggggtcattcaacagtt 391
Db      392 tatcagaacaacaggggggtcattcaacagctt 425

```

RESULT 9
AAA43182

ID AAA43182 standard; cDNA: 335 BP.

AC AAA43182;
XX

DT 21-AUG-2000 (first entry)
XX

DE Xenopus secreted expressed sequence tag SEQ ID NO:1922.
XX

KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytosolic; antibacterial; antifungal;
KW antiviral; antidiabetic; antitubercular; neurotrophic; antiparkinsonian;
KW antitumor; osteoprotective; neuroprotective; nootropic; antiparasitic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.

OS Xenopus sp.
XX

PN W0200021990-A1.
XX

PD 20-APR-2000.
XX

PF 15-OCT-1999; 99WO-US24205.
XX

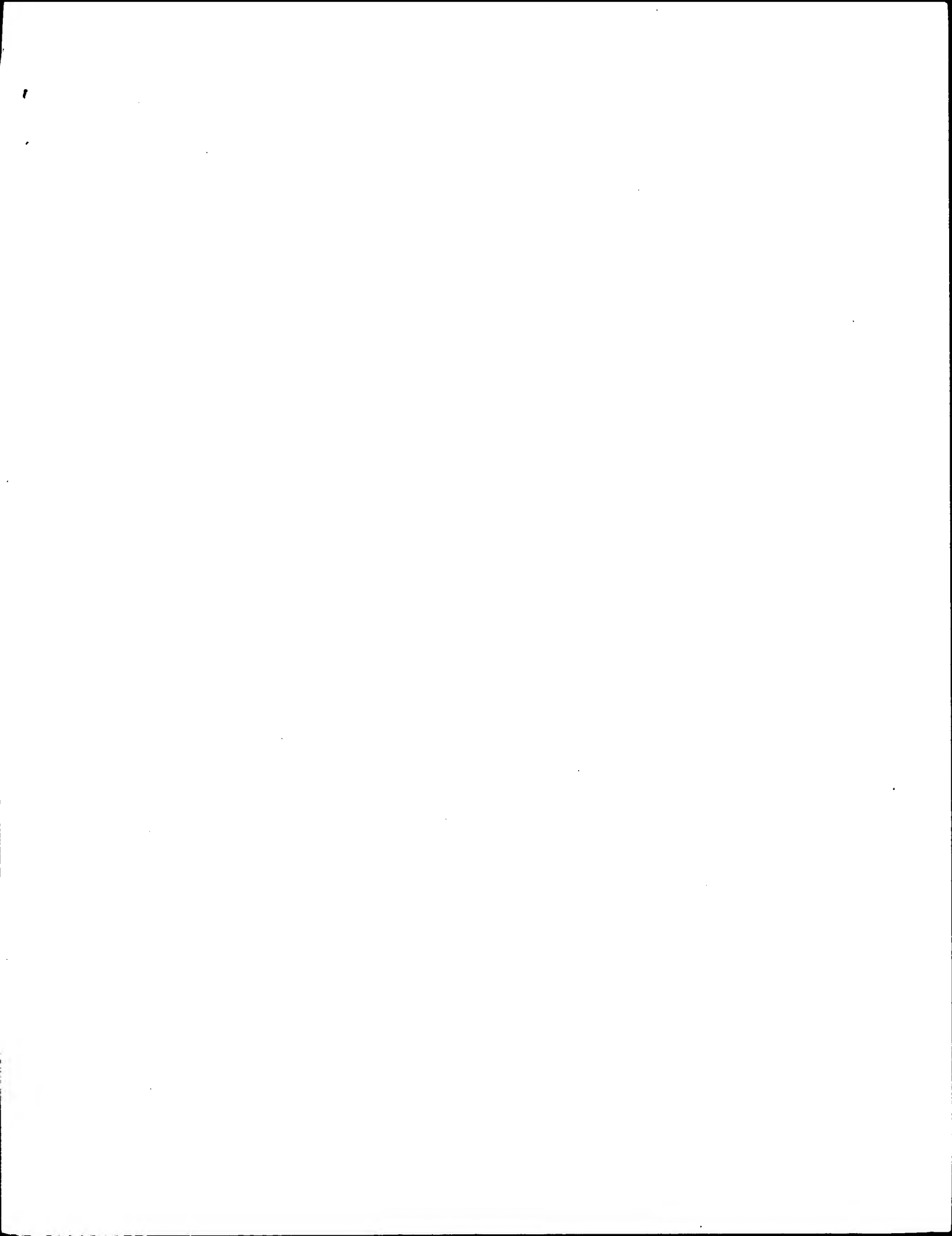
PR 15-OCT-1998; 98US-0104435.
XX

PA (GENY) GENETICS INST INC.
XX

Thu Aug 16 08:48:50 2001

us-09-484-964-1.rng

Page 15



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 15:19:49 ; Search time 86.75 seconds
(without alignments)
3197.012 Million cell updates/sec

Title: US-09-484-964-1
Perfect score: 1465
Sequence: 1 cgggaagattgttaacc.....gataatgcatgcacgc 1465

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1286.6	87.8	1514	2	US-09-213-768-1
2	77.6	5.3	473	2	US-08-853-974-2
3	77.6	5.3	473	3	US-09-172-988-2
4	50.6	3.5	7218	1	US-08-232-463-14
5	48.4	3.3	2110	4	US-09-419-459-1
6	46.6	3.2	19124	2	US-08-487-826B-13
7	46	3.1	7218	1	US-08-232-463-14
8	44.8	3.1	87350	3	US-08-781-891-79
9	44.4	3.0	734	3	US-09-014-583-1
10	44.4	3.0	1687	1	US-08-485-284A-1
11	44.2	3.0	1890	6	5312912-3
12	43.4	3.0	1441	4	US-08-821-994-63
13	42.6	2.9	6243	2	US-09-056-075-1
14	42	2.9	1104	4	US-08-307-499-57
15	42	2.9	1104	4	US-09-299-268-57
16	41.8	2.9	2435	4	US-09-306-593-1
17	41	2.8	1582	3	US-08-545-196B-10
18	41	2.8	1582	3	US-08-545-196B-12
19	40.4	2.8	1422	1	US-08-319-704-5
20	40.2	2.7	3933	2	US-08-731-722-3
21	40.2	2.7	3933	2	US-08-731-722-3
22	40	2.7	1198	3	US-09-248-335-27
23	40	2.7	1440	2	US-08-743-637B-174
24	40	2.7	1440	2	US-08-526-840B-174
25	39.8	2.7	2960	3	US-08-913-842-3
26	39.8	2.7	1046	1	US-08-361-467B-4
27	39.8	2.7	1046	1	US-08-484-332C-4

C	28	38.6	2.6	240	1	US-08-628-417-6	Sequence 6, Appli
C	29	38.6	2.6	1356	4	US-08-675-816-8	Sequence 8, Appli
C	30	38.6	2.6	2447	2	US-09-014-969-14	Sequence 14, Appli
C	31	38.6	2.6	2989	6	5378464-1	Patent No. 5378464
C	32	38.6	2.6	4032	1	US-08-107-748-3	Sequence 3, Appli
C	33	38.6	2.6	4032	1	US-08-245-809-4	Sequence 4, Appli
C	34	38.6	2.6	4032	5	PCR-US92-01385-3	Sequence 3, Appli
C	35	38.6	2.6	4358	4	US-08-675-816-7	Sequence 7, Appli
C	36	38.6	2.6	19124	2	US-08-487-826B-13	Sequence 13, Appli
C	37	38	2.6	209	1	US-08-146-421-3	Sequence 3, Appli
C	38	38	2.6	1147	1	US-08-665-716-1	Sequence 1, Appli
C	39	38	2.6	5718	3	US-08-714-918-48	Sequence 48, Appli
C	40	38	2.6	5718	4	US-09-265-315-48	Sequence 48, Appli
C	41	38	2.6	5718	4	US-09-265-315-48	Sequence 48, Appli
C	42	38	2.6	5718	4	US-09-266-417-48	Sequence 48, Appli
C	43	37.8	2.6	5852	1	US-07-867-106-2	Sequence 2, Appli
C	44	37.6	2.6	1183	2	US-08-731-722-8	Sequence 8, Appli
C	45	37.6	2.6	1332	2	US-09-057-762-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-213-768-1
Sequence 1, Application US/09213768
Patent No. 5985664
GENERAL INFORMATION:
APPLICANT: Brenda F. Baker
TITLE OF INVENTION: ANTISENSE MODULATION OF SENTRIN EXPRESSION
FILE REFERENCE: RTS-0026
CURRENT APPLICATION NUMBER: US/09/213,768
CURRENT FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 47
SEQ ID NO 1
LENGTH: 1514
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (136)..(441)
US-09-213-768-1

Query Match	87.8%	Score 1286.6;	DB 2;	Length 1514;
Best Local Similarity	94.1%	Pred. NO. 0;		
Matches 1384;	Conservative	0;	Mismatches 74;	Indels 13; Gaps 4;
QY	2	ggaagattgttaaccgagcgaaggttcgttaaccgagcgcgtctgtgcga	61	
DB	50	gggaagattgttaaccgagcgaaggttcgttaaccgagcgcgtctgtgcga	109	
QY	62	gaccccggtgaagcgcgtcatctgttcgaccgagcgaagcgaagcgaagcga	121	
DB	110	gaccccggtgaagcgcgtcatctgttcgaccgagcgaagcgaagcgaagcga	169	
QY	122	acttgaggataaagcgaaggttgatataataactcaagtcattgacagataga	181	
DB	170	acttgaggataaagcgaaggttgatataataactcaagtcattgacagataga	229	
QY	182	gtgagattcaactcaagtgaaatgacaacacatctcaagaactcaagaatcact	241	
DB	230	gtgagattcaactcaagtgaaatgacaacacatctcaagaactcaagaatcact	289	
QY	242	gttaagaagcgaaggttcctcaatgaattcaactcaggttccttgagggtcagaatg	301	
DB	290	gttaagaagcgaaggttcctcaatgaattcaactcaggttccttgagggtcagaatg	349	
QY	302	ctgataatcatctcaaaaagcctgggaatggaggaagaagtgtgattgattac	361	
DB	350	ctgataatcatctcaaaaagcctgggaatggaggaagaagtgtgattgattac	409	

TITLE OF INVENTION: RECOMBINANT FOVLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22313-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232,463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935,313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300, 6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: PTZp4r-F15
 US-08-232-463-14

Query Match	3.5%;	Score 50.6;	DB 1;	Length 7218;
Best Local Similarity	1.3%;	Pred. No. 0.0035;		
Matches	5;	Conservative 221;	Mismatches 145;	Indels 0;
			Gaps	0;

[illegible]

4
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFELINGER, F.
APPLICANT: FALKNER, F. G.

RESULT 5
US-09-419-459-1
; Sequence 1, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Cheng, Kuo-Juan
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2110
; TYPE: DNA
; ORGANISM: Pteromyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1749)
US-09-419-459-1

Query Match 3.3%; Score 48.4; DB 4; Length 2110;
Best Local Similarity 47.4%; Pred. No. 0.0073;
Matches 145; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 950 aaagcctttaaagattttttaaataatcttgacagaagaccagaagcttaattt 1009
DB 1720 aaatggtatggtatggtcgcgataatgtaataataagaataataaatttct 1779
QY 1010 catgaacataataagcgcatacgacagaatgatacaagaacactgctctttt 1069
DB 1780 aaagaaataatattttaaataataataataagaataatattatacaacatttcaata 1839
QY 1070 agatttatttgaatttggcctggagataggttttaaaagacatgctgtacca 1129
DB 1840 aaatgtaatttaaatatttatttatttatttattttaaataataataataagaaga 1899
QY 1130 gcttcataataataaataatcttgcaaaaacgcgaactgcttatttattt 1189
DB 1900 aaatataataataataataataataagaataataataatttatttatttactta 1959
QY 1190 gataagaagaaataatgcttaataagaagtttcttgacataactggaatggcsc 1249
DB 1960 aagcaaaaaaaagatttaataatacaagaatttttaagaatggaataatgtatttaa 2019
QY 1250 atggtta 1255
DB 2020 ataata 2025

RESULT 6
US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhuan
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 3.2%; Score 46.6; DB 2; Length 19124;
Best Local Similarity 49.4%; Pred. No. 0.049;
Matches 121; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 378 tcatcaacagttgagatcttcttatttcttcttccctcaatccttcttact 437
DB 15856 TTAATTAAATTAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTT 15797
QY 438 tttaaaatagcttcttcttgaatggtgctcaaaacgaatgaaacgcaccccat 497
DB 15796 TATGTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15737
QY 498 ctcttgaacatcgcggaattgaattcagtgctcaattatcattatggtgtttc 557
DB 15736 TTAATTAAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15677
QY 558 atgtgctgattttgtgtgacaaagccctcagtcctcatattacccctctttaa 617
DB 15676 AATGTTTTTTTTCTCTCTTTGTTTATTTTATTTTATTTTATTTTATTTTATTTATA 15617
QY 618 aaatt 622
DB 15616 AAATT 15612

RESULT 7
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA

Thu Aug 16 08:48:50 2001

us-09-484-964-1.rn1

Page 5

ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 3.1%; Score 46; DB 1; Length 7218;
Best Local Similarity 2.7%; Pred. No. 0.046; Indels 0; Gaps 0;
Matches 10; Conservative 212; Mismatches 152;

4 gaagattgttaaccccgagcaggtctcctaccgagccgctgctgagaga 63
1448 GAAGATTGTGTAACCCGAGCAGGTCTCCTACCAGCCGCTGCTGAGAGA 1389
64 ccccggtgagacccgcatcgtctgacgagagcaaacctcactgagac 123
1388 RRR 1329
124 ttggggataagagaggtgataatactcaagtcattgagagatagcag 183
1328 RRR 1269
184 gagattactcaagtgaatgacacacatcagaagaactcaagaatcact 243
1268 RRR 1209
244 caagagaggtgttcaatgactcactcaggttctcttgaaggtcagaat 303
1208 RRR 1149
304 gataatcactcaagaagactggaatgagagaagatgattgagtttcag 363
1148 RRR 1089
364 gaacaacgagggg 377
1088 RRR 1075

RESULT 8
US-08-781-891-79/C
Sequence 79, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui

APPLICANT: Yu, Chang-Eu
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 87350 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-79

Query Match 3.1%; Score 44.8; DB 3; Length 87350;
Best Local Similarity 5.1%; Pred. No. 0.25; Indels 1; Gaps 1;
Matches 108; Conservative 0; Mismatches 87;

182 gtgagattcactcaagtgaatgacacacatcctcaagaagaactcaat 241
12134 GTGAGATTCAAGTGAATGACACACGACCCGCTTAGTAACCAATGAAGCTTAT 12075
242 gtcaagagcaggtgttcaatgactcactcaggttctcttgaaggtcaga 301
12074 GTAAGGCGCACGGTTTGCCATGACGACGATGAGTTCAATTGACAGGCGCAATCA 12015
302 ctgataatcactcaagaagactggaatgagaggaagatgattgagttatc 361
12014 ATGAAACGACACACTTCGACATTGGAAGT-GAGCTTAGACAAACTGATGTCTCC 11956
362 aggaacaacgagggg 377
11955 AGCAGCAGACATGGAG 11940

RESULT 9
US-09-014-583-1
Sequence 1, Application US/09014583
Patent No. 6140077
GENERAL INFORMATION:
APPLICANT: NAKAMURA, Takeshi
APPLICANT: SUZUKI, Tadaoshi
APPLICANT: TOKUDA, Junko
APPLICANT: KATO, No. 6140077uo
APPLICANT: SAKAI, Yasuyoshi
APPLICANT: MOCHIZUKI, Daisuke
APPLICANT: TAKAHASHI, Hiroshi
TITLE OF INVENTION: METHOD FOR PRODUCING PHYTASE

GENERAL INFORMATION:
APPLICANT: SAKAI, YASUYOSHI
APPLICANT: TANI, YOSHIKI
APPLICANT: SHIBANO, YUJI
APPLICANT: KONDO, HIROTO
APPLICANT: HATANAKA, HARUO
TITLE OF INVENTION: NOVEL VECTOR HAVING PROMOTER THAT IS
TITLE OF INVENTION: INDUCIBLE BY METHANOL AND/OR GLYCEROL
NUMBER OF SEQUENCES: 18

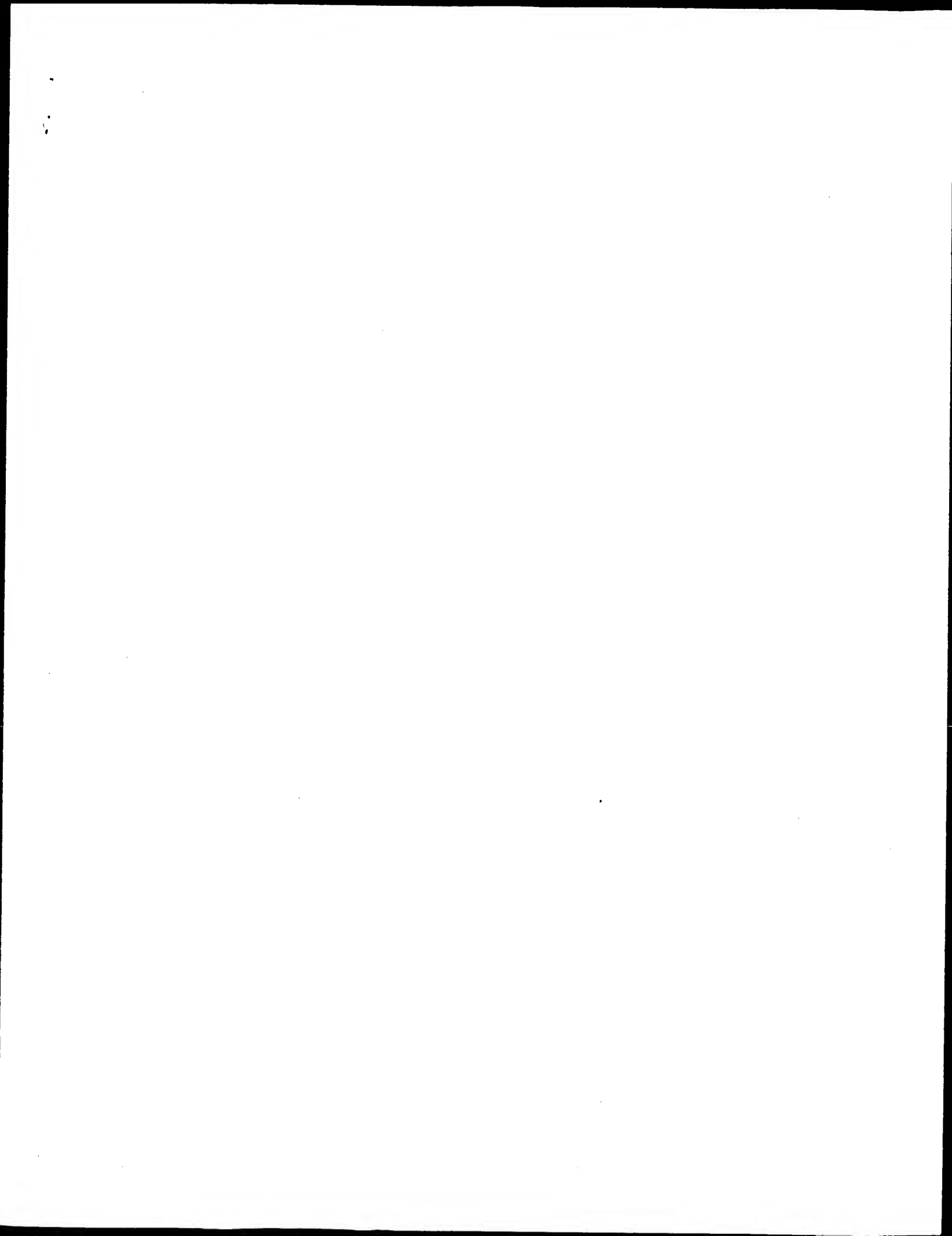
RESULT 10
US-08-485-284A-1
; Sequence 1, Application US/08485284A
; Patent No. 5750372

[illegible]

384 aacagtttagatattcttcttatcttttcctcccaatccttttaatttaa 443
 ||| ||| ||| ||| ||| ||| | ||||| |||||
 ||| ||| ||| ||| ||| ||| ||| ||||| |||||

Query Match	2.98;	Score 42;	DB 1;	Length 1104;
Best Local Similarity	47.78;	Pred. No.	0.21;	

SEQUENCE CHARACTERISTICS:
LENGTH: 1104 base pairs



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 15:18:29 ; Search time 1167.5 Seconds
(without alignments)
11861.611 Million cell updates/sec

Title: US-09-484-964-1

Perfect score: 1465
Sequence: 1 cgggaagatttgaacc.....gataaatggcattgcagcc 1465

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*


```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	945.2	64.5	1046	AL528799	AL528799 AL528799
2	933.2	63.7	998	AL563251	AL563251 AL563251
3	864.2	59.0	924	AL565221	AL565221 AL565221
4	852.8	58.2	868	AL521484	AL521484 AL521484
5	846.8	57.8	907	AL533666	AL533666 AL533666
6	843.8	57.6	913	AL583458	AL583458 AL583458
7	842.4	57.5	887	AL570877	AL570877 AL570877
8	840.6	57.4	898	AL544942	AL544942 AL544942
9	800.4	55.6	935	AL521483	AL521483 AL521483
10	779.6	53.2	933	AL583408	AL583408 AL583408
11	760.8	51.9	1175	AK002536	AK002536 Mus muscu
12	751.2	51.3	1180	AK011074	AK011074 Mus muscu
13	737	50.3	943	AL532393	AL532393 AL532393
14	734	50.1	1166	BG169405	BG169405 BG169405
15	732	50.0	804	BG540101	BG540101 BG540101
16	730	49.8	835	BG527341	BG527341 BG527341
17	724	49.4	822	BG527341	BG527341 BG527341
18	716	48.9	846	BG529400	BG529400 BG529400
19	714.4	48.8	827	AL532394	AL532394 AL532394
20	714.4	48.8	827	BG531082	BG531082 BG531082
21	714.2	48.6	798	AL538719	AL538719 AL538719
22	711.6	48.6	731	AL5066583	AL5066583 Ova47f05.x
23	702.2	47.9	979	BE738302	BE738302 BE738302
24	694.8	47.4	762	AL037282	AL037282 DKEP2564A
25	693.8	47.4	1045	BE887652	BE887652 BE887652
26	691.2	47.2	812	BE887652	BE887652 BE887652
27	690.2	47.1	786	BE887652	BE887652 BE887652
28	685	46.8	816	BE887652	BE887652 BE887652
29	684.2	46.7	858	BE887652	BE887652 BE887652
30	684	46.7	763	BE887652	BE887652 BE887652
31	683	46.6	781	BE887652	BE887652 BE887652
32	682.2	46.6	752	BE887652	BE887652 BE887652
33	678.6	46.3	761	BE887652	BE887652 BE887652
34	673.8	46.0	797	BE887652	BE887652 BE887652
35	672.4	45.9	858	BE887652	BE887652 BE887652
36	665.8	45.4	915	BE887652	BE887652 BE887652
37	664.8	45.4	733	BE887652	BE887652 BE887652
38	663	45.3	1148	BE887652	BE887652 BE887652
39	649	44.3	760	BE887652	BE887652 BE887652
40	645	44.0	769	BE887652	BE887652 BE887652
41	641.2	43.8	716	BE887652	BE887652 BE887652
42	639.6	43.7	716	BE887652	BE887652 BE887652
43	633	43.2	791	BE887652	BE887652 BE887652
44	627.8	42.9	770	BE887652	BE887652 BE887652
45	624	42.6	801	BE887652	BE887652 BE887652

ALIGNMENTS

Result No.	Score	Query Match	Length	DB ID	Description
1	945.2	64.5	1046	AL528799	AL528799 AL528799
2	933.2	63.7	998	AL563251	AL563251 AL563251
3	864.2	59.0	924	AL565221	AL565221 AL565221
4	852.8	58.2	868	AL521484	AL521484 AL521484
5	846.8	57.8	907	AL533666	AL533666 AL533666
6	843.8	57.6	913	AL583458	AL583458 AL583458
7	842.4	57.5	887	AL570877	AL570877 AL570877
8	840.6	57.4	898	AL544942	AL544942 AL544942
9	800.4	55.6	935	AL521483	AL521483 AL521483
10	779.6	53.2	933	AL583408	AL583408 AL583408
11	760.8	51.9	1175	AK002536	AK002536 Mus muscu
12	751.2	51.3	1180	AK011074	AK011074 Mus muscu
13	737	50.3	943	AL532393	AL532393 AL532393
14	734	50.1	1166	BG169405	BG169405 BG169405
15	732	50.0	804	BG540101	BG540101 BG540101
16	730	49.8	835	BG527341	BG527341 BG527341
17	724	49.4	822	BG527341	BG527341 BG527341
18	716	48.9	846	BG529400	BG529400 BG529400
19	714.4	48.8	827	AL532394	AL532394 AL532394
20	714.4	48.8	827	BG531082	BG531082 BG531082
21	714.2	48.6	798	AL538719	AL538719 AL538719
22	711.6	48.6	731	AL5066583	AL5066583 Ova47f05.x
23	702.2	47.9	979	BE738302	BE738302 BE738302
24	694.8	47.4	762	AL037282	AL037282 DKEP2564A
25	693.8	47.4	1045	BE887652	BE887652 BE887652
26	691.2	47.2	812	BE887652	BE887652 BE887652
27	690.2	47.1	786	BE887652	BE887652 BE887652
28	685	46.8	816	BE887652	BE887652 BE887652
29	684.2	46.7	858	BE887652	BE887652 BE887652
30	684	46.7	763	BE887652	BE887652 BE887652
31	683	46.6	781	BE887652	BE887652 BE887652
32	682.2	46.6	752	BE887652	BE887652 BE887652
33	678.6	46.3	761	BE887652	BE887652 BE887652
34	673.8	46.0	797	BE887652	BE887652 BE887652
35	672.4	45.9	858	BE887652	BE887652 BE887652
36	665.8	45.4	915	BE887652	BE887652 BE887652
37	664.8	45.4	733	BE887652	BE887652 BE887652
38	663	45.3	1148	BE887652	BE887652 BE887652
39	649	44.3	760	BE887652	BE887652 BE887652
40	645	44.0	769	BE887652	BE887652 BE887652
41	641.2	43.8	716	BE887652	BE887652 BE887652
42	639.6	43.7	716	BE887652	BE887652 BE887652
43	633	43.2	791	BE887652	BE887652 BE887652
44	627.8	42.9	770	BE887652	BE887652 BE887652
45	624	42.6	801	BE887652	BE887652 BE887652

FEATURES

Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr.
Location/Qualifiers

source

1. 1046

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CSDD001YH20"

/clone_11b="LTI_NFL001_NBC4"

/sex="male"

/tissue="neuroblastoma cells"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@life.com URL: http://fulllength.invitrogen.com"

BASE COUNT 310 a 185 c 231 g 318 t 2 others

ORIGIN

Query Match 64.5% Score 945.2; DB 106; Length 1046;

Best Local Similarity 98.7%; Pred. No. 2e-189;

Matches 983; Conservative 1; Mismatches 9; Indels 3; Gaps 3;

2 ggaagagatttgaaccccgagagaggttcttaccgagcgctgtcgga 61

33 gagaagatttgaaccccgagagaggttcttaccgagcgctgtcgga 92

62 gaccccgagagagagagagagagagagagagagagagagagagag 121

93 gaccccgagagagagagagagagagagagagagagagagagagag 152

122 acttgag 181

153 acttgag 212

182 gtag 241

213 gtag 272

242 gtag 301

273 gtag 332

302 gtag 361

333 gtag 392

362 gtag 421

393 gtag 452

422 gtag 481

453 gtag 512

482 gtag 541

513 gtag 572

542 gtag 601

573 gtag 632

602 gtag 661

633 gtag 692

RESULT 1
AL528799 1046 bp mRNA EST 13-FEB-2001
LOCUS AL528799 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSDD001YH20 5
DEFINITION Prime, mRNA sequence.
ACCESSION AL528799
VERSION AL528799.1 GI:12792292
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1046)
AUTHORS L.L.W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Query Match 64.5% Score 945.2; DB 106; Length 1046;
Best Local Similarity 98.7%; Pred. No. 2e-189;
Matches 983; Conservative 1; Mismatches 9; Indels 3; Gaps 3;
2 ggaagagatttgaaccccgagagaggttcttaccgagcgctgtcgga 61
33 gagaagatttgaaccccgagagaggttcttaccgagcgctgtcgga 92
62 gaccccgagagagagagagagagagagagagagagagagagagag 121
93 gaccccgagagagagagagagagagagagagagagagagagagag 152
122 acttgag 181
153 acttgag 212
182 gtag 241
213 gtag 272
242 gtag 301
273 gtag 332
302 gtag 361
333 gtag 392
362 gtag 421
393 gtag 452
422 gtag 481
453 gtag 512
482 gtag 541
513 gtag 572
542 gtag 601
573 gtag 632
602 gtag 661
633 gtag 692

QY 662 ttctcagctgtggtgataaataagatcagcaatgcaagctgtcattatgactttcca 721
 Db 693 TTTTCAGGCTTGTGTGATTAATTAAGATGACCAATGCAAGTGTCATATATACCTTCCA 752
 QY 722 attgacctgactgtc-agaatgacttacttacttacttacttacttacttacttactt 780
 Db 753 ATTGGCCCTGATGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 812
 QY 781 gatgaagcttttcaagaacgactgtggaataatgaccttcccttacttacttacttactt 840
 Db 813 GATGGAATGTTTTCAGAAACGAACTGGAATAATGACCTTCCCTTAACCTGGAAGCTA 872
 QY 841 ctttaaaatttgaggtcgtgagcaaaagaagaatgataatgacttgaagcaagatga 900
 Db 873 CTTTAAATTTGAGGCTGCTGAGCAAAAGAAAGAAATATGAGTTGAGTCAAGATGA 932
 QY 901 cagataagtgagagtaataactaactcacaagatgcttacttgaagaaagcatatt 960
 Db 933 CAGATAAGGTGAGATGATGACTTACCTCAAAAGATGAGC-TYACTGAAAGAAAGCATTTT 991
 QY 961 aagatttttaaaatctgtcagaagaatcccaaga 996
 Db 992 AAGATTTTTT-AAATCTTGTCAAGKACCAAGAGA 1026

RESULT 2
 AL563251/c 998 bp mRNA EST 16-FEB-2001
 LOCUS AL563251 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD001YH20 3
 DEFINITION prime, mRNA sequence.
 ACCESSION AL563251
 VERSION AL563251.1 GI:12912463
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 998)
 AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 1. 998
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="LTI_NFL001_NBC4"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library of the pCMVSPORT 6
 by Life Technologies. Contact: Feng Liang Life
 Technologies, a division of Invitrogen 9800 Medical Center
 Drive Rockville, Maryland 20850, USA Fax: (41) 301 610
 8371 Email: fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 336 a 193 c 160 g 298 t 11 others
 ORIGIN

Query Match 63.7%; Score 933.2; DB 106; Length 998;
 Best Local Similarity 96.5%; Pred. No. 6.9e-187;
 Matches 964; Conservative 8; Mismatches 25; Indels 2; Gaps 2;

QY 152 ttaactcaagatcatttgcagcgatagcgatgacttcaaatgaaatgacaa 211
 Db 998 TTAACCTCAAAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 939
 QY 212 cacatctcaagaacatcaagaatcactgtcagaagaagggtgttccaatgaaatc 271
 Db 938 CACTCTCAAGAACTCAAAAGATCAAACTGCAAAAGACAGGCTGTTCAATGAATTAC 879
 QY 272 tcaagttctcttgaagggtcagaagaatgctgataatcaatcccaagaactgga 331
 Db 878 TCAGGTTTCTTTGAGGCTGAGAGATGCTATATATCAATCACTCAAAAGACAGGGA 819
 QY 332 tggaggaagaatgattgattgaagtttcaagaaacagggggtcattcaaatg 391
 Db 818 TGAGAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 759
 QY 392 agatacttttatt 451
 Db 758 AGATATCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 699
 QY 452 ttgtgaatgtggtgttcaaaaaggaaattgaagaactggcaaccacatcttgaacac 511
 Db 698 TTTGTATNTGTGTCTTCAAAAGCAATTTGAAGTGGACCCCATCTCTTTGAACNTC 639
 QY 512 tggtaattggaattcagtgctcattatcattatgattgttttcatgtgtgattt 571
 Db 638 TGCTTATTTGATTTCTAGTGCTCATTTATTCATTTATTTGTTTTCATTTGCTGATTTT 579
 QY 572 tggatcaagaactcagtcctccttcaatataccctccttctttaaataaagtgatga 631
 Db 578 TGCTGCTCAAGCTCAGTCCCTTCATATTTACCTCTCTTTTAAAAATTAACGTGTGA 519
 QY 632 cagaagagtcacatttttcaagaatgacatttcaagctgtgtgtataaataagtcg 691
 Db 518 CAGAGGGTCACTTTTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATG 459
 QY 692 accaatgcaagtgttcaataagatcacttcaattgagcctgtgttc-agaatgattac 750
 Db 458 ACCAATGCAAGTGTTCATTAATGATGATGATGATGATGATGATGATGATGATGATG 399
 QY 751 ttactcctgagctgtgacttcaatgagtgagagtggaagttttcaagaagaactgactg 810
 Db 398 TTTACTCCTGAGCTGTACATTTTCACTGAGATGAGATGAGATGAGATGAGATGAGATG 339
 QY 811 gaaaatgacatttcccttcaatgaaacttcaattttaaatttgaggggtctggaccaaaag 870
 Db 338 GAAATGACCTTTCCTTTACTTGAAGCTTAAATTTGAGGCTTGAGACCAAG 279
 QY 871 aagaagataatcaggttgaagtcagaatgacagataagtgagagtaataactca 930
 Db 278 GACAGGAATATCAGGTTGAGATGAAGTAAAGATTAAGTAAAGTAAATTAATCACTCA 219
 QY 931 aagatgcttcactgaaagaagacattttaaagatttttaaaactctgtcagaagtc 990
 Db 218 AAGATGCTTTACTGAGAAAGAAAGCAATTTAAAGATTTTAAAAATCTTTCAGAAATC 159
 QY 991 ccgaagaagttcaatttcaatgaatcaatgaatgaatgaatgaatgaatgaatgaatga 1050
 Db 158 CCGAAAGATTTCAATTTTCAATTTGCAATTAATTAATTAATTAATTAATTAATTAATTA 99
 QY 1051 acgaagaactgctctttagatttatttgaatttgaatttgaatttgaatttgaattt 1110
 Db 98 ACAGAAAGATGCTCTTTT-GATTTATTTGATCTTTTGGCTGGAAATGAGGTTTAA 40
 QY 1111 atggaactgtcgtgacagctcatttaataataacaat 1149
 Db 39 ATGACATTTGCTGTACCACTTCATTAATAAACAAT 1

RESULT 3
 AL565221/c 924 bp mRNA EST 16-FEB-2001
 LOCUS AL565221 LTI_FL013_Fbrn1 Homo sapiens cDNA clone CS0DF002Y003 3
 DEFINITION


```

Db      464  |||||||
          TCAATCCTTTTATTTTAAATAAGTCTTTTAAAGTGTTGCAAAACGGAATTG 523
Qy      482  aaaaatggaaccacatctcttgaacatctgtaattgaattcgaatgctcattatc 541
          |||||||
Db      524  AAACATGACACCCATCTCTTGAACATCTGTATTTGAATTTAGTGCATATTTC 583
          |||||||
Qy      542  attatgttcttctcattatgctgctatttggatgatacgaagccctgaatcatt 601
          |||||||
Db      584  ATATATGTTGTTTCATGCTGCTGATTTTGTGATCAAGCCCTAGTCCCTCATATT 643
          |||||||
Qy      602  accctctctctttaaataatcagtgatgacagagagcacccttttgaagaaattga 661
          |||||||
Db      644  ACCCTCTCTTTTAAATAATAGTGTGACAGAGAGTCCCTTTTCAAGACATGCA 703
          |||||||
Qy      662  ttctcagcttgatgataaataagatcgaacatgacagtgctcacaattcattcca 721
          |||||||
Db      704  TTTTCAGGCTTGCTGTGATTAATAGATGACCAATGCAAGTTTCATATGACTTTCCA 763
          |||||||
Qy      722  attgacctgatagtc-agaatgtgattactcctcctgagctgacttccagtggga 780
          |||||||
Db      764  ATGGCCCTGATGTTCTACATGTGATTAATCACTCTGAGCTGACTTTCAGTGGGA 823
          |||||||
Qy      781  gatgagaatttctcagaagacgaactgtggaataatgaccttctcattgaagcga 840
          |||||||
Db      824  GATGGAAGTTTTCAGAGAACTGAATGGAATAATGACCTTCTTAACCTTGAAGCTA 883
          |||||||
Qy      841  ctcttaaatattgaggtctgac 864
          |||||||
Db      884  CTTTAAATTTGAGGGCGCTGRC 907
          |||||||

```

```

RESULT 6
AL583458 913 bp mRNA EST 16-FEB-2001
LOCUS AL583458 LTI.NF1010.BC2 Homo sapiens cDNA clone CS0D1012YD05 5
DEFINITION prime, mRNA sequence.
ACCESSION AL583458
VERSION AL583458.1 GI:12952440
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 913)
          Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
          Full-length cDNA libraries and normalization
          unpublished (2001)
COMMENT Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 Evry cedex - France
          Email: sequef@genoscope.cns.fr, web : www.genoscope.cns.fr.
          Location/Qualifiers
            1. 913
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="CS0D1012YD05"
              /clone_1b="LTI.NF1010.BC2"
              /sex="male"
              /tissue-type="B cells from Burkitt lymphoma"
              /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
              was primed with a NotI-oligo(dT) primer. Five prime end
              enriched, double-stranded cDNA was digested with NotI and
              cloned into the NotI and EcoRV sites of the pCMVSPORT 6
              vector. Library was normalized. Library was constructed by
              Life Technologies. Contact: Feng Liang Life Technologies,
              a division of Invitrogen 9800 Medical Center Drive
              Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
              Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com/ 1 others

```

```

BASE COUNT 268 a 166 c 199 g 279 t
ORIGIN

```

```

Query Match 57.6% Score 843.8 DB 106 Length 913;
Best Local Similarity 98.9% Pred. No. 5e-168 7; Indels 3; Gaps 2;
Matches 871; Conservative 0; Mismatches

```

```

Qy      15  aaaccggagcggaggtctgcttaccgaagccgctgctgctgagaccccggtga 74
          |||||||
Db      1  AAACCCCGGACGAGGTTCTCTCTACCGGAGCGCGCTGTGGAGACCCCGGGTGA 60
          |||||||
Qy      75  agccacgctcatcaltgctgaccagaggaacacacacacacacacacacacacac 134
          |||||||
Db      61  AGCCACCGTCAATGCTGCTGACAGAGGCAAAACCTTCAACTAGAGACTTGGGGATAA 120
          |||||||
Qy      135  gaagagaagtgaatatataaactcaaaagatgacagatagacagatgacatcact 194
          |||||||
Db      121  GAAGGAAGGTGAATATATTAATCAAGTCAAGTATGACAGATGACAGATGATTCATT 180
          |||||||
Qy      195  caaagtgaatgaacaacacacacacacacacacacacacacacacacacacacag 254
          |||||||
Db      161  CAAATGAAATGACAAACACATCTCAAGAACTCAAAATCATATCTGTCAAAGACAGG 240
          |||||||
Qy      255  tttcacaatgaatcactcaggttctccttggaggctcagaagattgctgatactac 314
          |||||||
Db      241  TGTTCATGAAATTCACACAGGTTCTCTTGAAGGCTCAGAAATGCTGATATATCATAC 300
          |||||||
Qy      315  tccaaaagaactgggaatggaggaagatgagatgagatgagatgagatgagatgag 374
          |||||||
Db      301  TCCAAAGAACTGGGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
          |||||||
Qy      375  ggtcattcaacagtttgaatattcttcttcttcttcttcttcttcttcttcttct 434
          |||||||
Db      361  GGGTATTCACAGTTTATGATATCTTTTATTTTATTTTATTTTATTTTATTTTATTTT 420
          |||||||
Qy      435  attttaaataatgcttcttcttcttcttcttcttcttcttcttcttcttcttct 494
          |||||||
Db      421  ATTTTAAATAATGAGTCTTTTGTGATGAGGTTTCAAAACGGAATGAAACCTGGACCC 480
          |||||||
Qy      495  catcctcttgaacaacatcgtgaatggaatggaatggaatggaatggaatggaatg 554
          |||||||
Db      481  CATCTCTTGAACATCTGCTGATTAATGATTAATGATTAATGATTAATGATTAATGAT 540
          |||||||
Qy      555  ttcattgtgctgatttcttcttcttcttcttcttcttcttcttcttcttcttct 614
          |||||||
Db      541  TTCATTTGCTGATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
          |||||||
Qy      615  taaaataatcgtgctgacagagaagtgacacacacacacacacacacacacacac 674
          |||||||
Db      601  TAAAAATTAAGTGTGACAGAGAGAGTACCTTTTTCAGGACATTCATTTTCAGGCTGT 660
          |||||||
Qy      675  gttgataaataagatgacacacacacacacacacacacacacacacacacacacac 734
          |||||||
Db      661  GGTGATTAATAATGATGACCAATGCAAGTGTTCATTAATGATGATGATGATGATGATG 720
          |||||||
Qy      735  ttc-agaatgattactcactcctgagctgacttgaacttgaatgagatgagatgag 793
          |||||||
Db      721  TTCTGACATGATTAATCTTCACTCCGAGCTGACTTCACTGAGAGATGAGAAATTTT 780
          |||||||
Qy      794  cagagaactgaactgtgaaataatgacacacacacacacacacacacacacacac 853
          |||||||
Db      781  CAGAGAACTGAACTGTGGAATAATGACCTTTTCACTTGAAGTCACTTTTAAATTTG 840
          |||||||
Qy      854  aggtctgacacaaagaagaatcaggttgaagca 894
          |||||||
Db      841  AGGCTCTGAC-AAAAGAAAGATATCAGGTGAAGTCA 879
          |||||||

```

```

RESULT 7
AL570877 887 bp mRNA EST 16-FEB-2001
LOCUS AL570877 LTI.NF1006.PL2 Homo sapiens cDNA clone CS0D1012YA10 3
DEFINITION prime, mRNA sequence.
ACCESSION AL570877
VERSION AL570877.1 GI:12927614

```



```

Db      9  GAGAAAGATTTTAAACCCCGAGCGAGTCTGCTACCCGAGGCCCTGCTGCGCA 68
Qy      62  gacccccgggtaagccaccgtcatcagtcgacaggaagcaaaccttcaactgagg 121
Db      69  GACCCCGGGGTGAACCCACCGTCATCATGTCTGACAGGAGGCAAAACCTTCACTGAGG 128
Qy      122  acttgggggataagaagcaagtgatattataactaaagtcatttgacagggtaaca 181
Db      129  ACTTGGGGGATTAAGAGGAAGGTGATATATATTAATTAACCTCAAGTCTGAGCGATACCA 188
Qy      182  gtgagattcactccaagtgaataatgacacacacacccaagaactccaagaatcact 241
Db      189  GTGAGATTCTACTTCAAGAGTAATGACACACACTCTCAAGAACTCAAGAGATCATACT 248
Qy      242  gtcaaaagcaagggtgttccaatgaatcaactcaagttcttcttgaagggtcagagaat 301
Db      249  GTCAAAGACAGGGGTCTCCATGATGATTCACCTGAGTTCTCTTGAAGGTGAGAGAAATGG 308
Qy      302  ctgataatcatctcccaaaagacatgggaatgaggaagaagaatgattgattatc 361
Db      309  CTGATATATATCTCCAAAGAACTGGGAATGAGAGAAAGATGTGATTAAGTTATTC 368
Qy      362  aggaacaacagggggtgtcacaacatgtagatacttcttatttcttcttcc 421
Db      369  AGGACAAACGGGGGTCTCATTCACAGTTTATATATCTTTTATTTTCTTTTCC 428
Qy      422  tcaatccttttattttaaataatgctcttcttgaatgagtggtgttcaaaaggaatg 481
Db      429  TCAATCTTTTATTTTAAATAGTTCTTTGTAATGTGTTGTTCAAAACGGAATGG 488
Qy      482  aaaaacgtgaccccatcttcttgaacacatgtaattgaattgaagtcattatc 541
Db      489  AAAACGTGACCCCATCTTTGAAACATGTGTATTAATGAAATTCAGTGTCAATATTC 548
Qy      542  attatgttcttcttcatgctgctgatttcttgaatgaacacccatccatc 601
Db      549  ATTATGTGTGTTTTCATTTGCTGATTTTGTGTATTCAGGCTCAGTCCCTCATATTT 608
Qy      602  acccttccctttaaataatcagtgacagagagtgacaccccttcaagacattgca 661
Db      609  ACCCTCTCTTTTAAATAATAGTGTGACAGAGAGTACCTTTTTCAGACATTTGCA 668
Qy      662  ttttggagcttgggtgataaataagatgcacacgaatggttataatgacttca 721
Db      669  TTTTCAGGCTGTGGTGAATAATAGATGACCAATGCAAGTGTATATAGCTTTTCCA 728
Qy      722  attggccctgagtc-agaatgtagtacttactccttgaactgtagcttcaagtggga 780
Db      729  ATTGGCCCTGATGTTCTAGATGTGATTAATTAATTAATTAATTAATTAATTAATTAAT 788
Qy      781  gatgaaagtttctcagagaactgactgtggaataatgaccccttcc-ttaacttgaagt 839
Db      789  GATGGAAGTGTTCAGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 848
Qy      840  acctttaaataatgagtgctgacccaagaagaagaagaatcaggttga 890
Db      849  ACTTTTAAATTTGAGGGCTGAGAC-AAARAAGAGATATAGGTTGGA 898

```

```

RESULT 9
AL521483 935 bp mRNA EST 13-FEB-2001
LOCUS AL521483.L1_NFL004_NBC2 Homo sapiens cDNA clone CS0DB001YN17 3
DEFINITION prime, mRNA sequence.
ACCESSION AL521483
VERSION AL521483.1 GI:12784976
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 935)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

```

```

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
1..935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB001YN17"
/clone_id="L1_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : filiang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 254 a 182 c 186 g 286 t 27 others
ORIGIN
Query Match 54.6%; Score 800.4; DB 105; Length 935;
Best Local Similarity 92.2%; Pred. No. 7.2e-159;
Matches 864; Conservative 14; Mismatches 55; Indels 4; Gaps 4;
Qy 216 tctcaagaactcaagaatcactcgtcaagacaggggtgttccatgattcactcag 275
Db 935 TCTCAAGAAAGTCAAGAAATCATTTGTCAAAGAGGTTGTCATGATTCACAG 876
Qy 276 gttctcttgaagggtcagaagattgctgataatcactcacaagaactggaatgga 335
Db 875 GTTTCCTTTGAGGGTCAAGAAATGCTGATTAATCAATCTCAAAAGAACTGGGAATGGA 816
Qy 336 gaaagaagatgtagtgaattatcagaacaaacgggggtgattcaacagtttagat 395
Db 815 GGAAGAAGATGATGATGAATGATGATGATGATGATGATGATGATGATGATGATGAT 757
Qy 396 attcttttatttttcttcttccctcaatcctttttat-ttttaaatatttctt 454
Db 756 ATTCTTTTATTTTTCCTTCCCTCAATCTSTTTATCGCTCAAAATATGCTCTTNT 697
Qy 455 tgaatgtggtgttcaaaaggaattgaaacatgaccccatctcttgaacatgag 514
Db 696 TGNAAAGNGGTGTCNAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA 637
Qy 515 taattgaattcagtgctcattatcattatgttgttttcaatggtgctgattttgg 574
Db 636 TAATTTGAATTTAGTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 577
Qy 575 tgataagcctagtccttcaattaccccttccctttaaataatgagtgagacag 634
Db 576 TGACCAAGCTCAGCCCTCATATTAACCCCTCTGTAATAATATGAGTGTGACAG 517
Qy 635 aaggttcaacttctcagacatcatttcaagctgtgtgtaataatgacgcagc 694
Db 516 AAGAGTCAACCTTTTCAGAGACATGCAATTTTCAAGCTGTGTGATTAATTAATTAAT 457
Qy 695 aatgcaagtgatcattgaacttccaatgtgcccctgagtc-agaatgtagtacttc 753
Db 456 AATGCAAGTGTATTAATGACCTCCCAATTTGCCCCGAGTGTTCAGATTAATTAATTC 397
Qy 754 actcctgagctgagcttcaagtgagagatggaatttcaagaagaactgagtgaa 813
Db 396 ACCCCGAGCTGTGACTTTCAGTGGAGATGGAAGTTTTCAGAGAACTGAACGTGTGAA 337

```

QY	814	aaatgaaccttccctaaacttgaagctactctttaaattgaaggtctggccccaaaagaag	873
Db	336	AAATGACCTTTCCCTAAACCTGAACCTACTBTATAAATTGAGGCTCTGGACCAAAAGAG	2777
QY	874	aggaatatcatgaattggaagtcagaatgtacagatagaatgtgagaatgataagtactccaag	9333
Db	276	AGGAATATACAGCTTTGAAGTCAAGATGACAGTAAGTGGAAGATTAATACACTCCAAAG	217
QY	934	atggcttcaactcgaagaaaggcatctttaagatttctttaaactctgtccagaagatccca	9933
Db	216	ATGGCTTCACCTGAGAAAGGACATKTTTAAGATKTTTAAAAATCTTGTCAGGAAGACCCA	157
QY	994	gaaaagtctcaatttcttaagacatataaagaatctacatcagagaatagtatacaaca	105
Db	156	GAAAGTTCTATTTTCAATTATGCAATTATTAATAAAGCTACATGACAGAAATGAACACACA	97
QY	1054	gaacactgtctcttattagattatttgtaacttlttgccctggatattggtttaaattg	111
Db	96	GAAACACATGC -CCCTTTTGAGATGCVATNTNGTACSGNNINGCCTGGGAAATGSGCCCAATG	38
QY	1114	gaacattgtctgacagcttcaatataaataacaata	1150
Db	37	GACATGTCTGTACACACTTCATTATAAATAAACATA	1

RESULT	10
AL583408/c	
LOCUS	
DEFINITION	AL583408 933 bp mRNA EST 16-FEB-2001
ACCESSION	AL583408 L11_NF010_Bc2 Homo sapiens cDNA clone CS0DL012YD05 3
VERSION	AL583408
KEYWORDS	AL583408.1 GI:12952341
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL	1 (bases 1 to 933)
COMMENT	L1.W.B., Gruber,C., Jesse,J. and Polyes,D.
	Full-length cDNA libraries and normalization
	Unpublished (2001)
	Contact: Genoscope
	Genoscope - Centre National de Sequencage
	BP 191 91006 EVRY cedex - France
FEATURES	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
source	Location/Qualifiers
	1..933

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL012YD05"
/clone_lib="LTI_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies a division of Invitrogen 9600 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

```

Query Match	53.28;	Score 779.6;	DB 106;	Length 933;
Best Local Similarity	95.38;	Pred. No. 1.8e-154		
Matches 871;	Conservative	23;	Indels 12;	Gaps 7;
502	tttaaacatcgcggaattgattctcagtgcctcatattatcattatgttggcttcacatg	561		

Db	922	TTGAMACATCTGGATTGGAATCTAGGCTCAATAACTT-----ANGTTTGCTTCATTG	868
OY	562	tgctgattcttgtygatcaagcgcacgtcccttcacatcattaccctcctcttcttaaaat	621
Db	867	TGCTGA-TTTTGGATCAACGCTCAGGCCCTTCATATTACCTCCCTTTTAAAAAT	809
OY	622	tactgtgcaagagaggttcaaccttttcaggacattgcatttcaagcttgtgtgata	681
Db	808	TACGGTGCACAGAGAGGTCACTTTTTCAGGACATTGCATTTTCAGGCTTGGTGGAAT	749
OY	682	aataagatcgacccaatgcaagtgctcaataagcttccatcttcaggccctgaagtc-aa	740
Db	748	AATTAAGATCGACCAATGCAAGTGTCTCAATTAAGACTTTCCATTGGCCCTGATGTTTAC	689
OY	741	atgtagtacttcaactccttgagctgttacttcagtgagatggaagtttccagagaa	800
Db	688	ATGCGATTACTTCACTCCTGGACTGTGACTTTTCAGTGGAGATGGAAGTTTTCAGAGA	629
OY	801	ctgaactgtggaaaaaagaccttcccttaacttgaagctactctttaaattgaggtct	860
Db	628	CTGAACCTGTGAAAAAAGACCTTTCCTTAAGCTGGAAGCTTCCTTTAAATTTGACGGCTC	569
OY	861	ggacccaaaagagggagatcacaggttgaagtaagaatgcacgaataagtgagagtaatg	920
Db	568	GGACCCAAAAGAGGGAATATCAGGTTGAAGTCAAGTGCACATTAAGGTGAAGTAATG	509
OY	921	actactccaaaagaatgcttcaactcgaagaagaagcatttlaagaatttctlaaaactctg	980
Db	508	ACTTAATCTCAAAAGATGGCTTCAGTGAAGAAAGCAATTTTAAGATTTTAAAAATCTTG	449
OY	981	tcaagagatcccaagaaaagtcttaatttcaatagcaatataaagaatatacatgcaga	1044
Db	448	TCAAGAGATCCCAAAAAAGTTCTAATTTTCATTAGCAATTAATAAAGCATACATGCACA	389
OY	1041	aagcaatacaacagaaacacgcctcttttgaatttatttgaactcttttggccctggata	1100
Db	388	AATTAATATCACACAGACACACTGCTCTTTT-GATTTTATTTGTACTTTTGGCTGGGATA	330
OY	1101	tgaggtttaaattggacattgctgtccacagctcattcaataataacaatctgtcaaa	1160
Db	329	TGGSTTTAATATGACATGTCTGTACCACTTCATTAATAATTAACAATATTTGT-AAAA	271
OY	1161	atcgtaactaatgcttatttattttaaattgatatagaagaagaaaaaatgcctaaataag	1220
Db	270	ATCATACTAATGCTTATTATTATTATTTAATGTATAGAAACAAAAAAGCCCTAAATTAAG	211
OY	1221	ttttctgcataaactgtaaatgtcaatgttcaaaatltttcttataactgcgtaca	1280
Db	210	TTCTCTTGCAATTAATATCTGGAATATGCACATGGTACAAATTTTTTCTTCATTACGTGACA	151
OY	1281	ggagatgcttaatgactctg--gagcaactggaagttactgaagtgccctcgaatacaag	1336
Db	150	GTAATGATGTTAATGACTCTGAGACACACTGAAGTACTGAAGTCCCTTBGANTCAAG	91
OY	1339	gatttaataaggccaonaatacctttaaatactcagtgctcgttltttttaaanaact	1398
Db	90	GATTTAATTAAAGCVCACAAATACCTTTTAAATCTAGTGTCTG-TTTTTAAAMACTT	32
OY	1399	gatatcccgatg 1412	
Db	31	GATATTCTCTGTATG 18	

RESULT 11					
AK002536					
LOCUS	AK002536	1175 bp	mRNA	HTC	08-FEB-2001
DEFINITION	Mus musculus adult male kidney cDNA, RIKEN full-length enriched				
ACCESSION	AK002536	library, clone:061001f08, full insert sequence.			
VERSION	AK002536.1	GI:12832590			
KEYWORDS	CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library				

ORGANISM	Mus musculus
STRAIN: 0010011100	

clone:0610011F08.
Mus musculus

[illegible]

AUTHORS Carninci, P. and Hayashizaki, Y.

JOURNAL
Methods Enzymol. 303, 19-44 (1999)

High-efficiency full-length cDNA cloning
Methods Enzymol. 303, 19-44 (1999)

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sug

TITLE

Itch, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to


```

|||||
Db 88 GACCCCGGGTAAATCCA-CGTACCATGTCTGACACGAGGAGCAAAACCTTCAACTGAG 146
Oy 122 acttgggggataagaagaagtgatataataaactaaagtccttgacaagataaga 181
Db 147 ACTTAGGCGATTAAGAGAGAGAGAAATACATTAACCTAAATTTTGTGACGAGTACCA 206
Oy 182 gtgagattcctaagaatgaaatgaacacacatcccaagaactcaagaatcact 241
Db 207 GTGAGATACATTCAAGATGAAATGACACACATCTCAAGAAATCTCAAGATATATCT 266
Oy 242 gtcaagaacagggtgttcccaatgaatcactcaagttctcttgagggtcagaagt 301
Db 267 GTCAAGAGACAGGAGATCCATGAATTCACAGGTTCTCTTGAAGTCCAGAAATTTG 326
Oy 302 ctgataatacattcccaagaactcgtgagatgaggaagaagatgattgagattc 361
Db 327 CTGATATATCTACTCGAAGAACTGGGATGAGAGAGAGATGATGATGAAGTTATC 386
Oy 362 aggaacaacagggtgtcattcaacagttatgata-----tctcttattttttt 416
Db 387 AGAACAACAGGGGGGCTCAGCTGACGTTAGATATCTTTTATTTTATTTTCT 446
Oy 417 ttcctcaatcccttttattttaaataagttcttcttgatgtgtgttcaaacg- 475
Db 447 TTCCCTCAATCTTTTATTTTATTTTAAATAGTCTTTTGTATGTGTGTCATAAATA 506
Oy 476 gaattgaacatcgacacccatctcttgaacaatcgtgaattgaattcattgctca 535
Db 507 AAATGAAATCGACCTGACCTCATCTTTAGAA-----CATATGAAATTCATGCTTCA 557
Oy 536 ttattcaattgttctgttcttcttctgtcatttctgtatcaagctcagctccct 595
Db 558 AATATCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 617
Oy 596 cataatccctcctctttaa---aataagtgagacagagaaggtcacttctcag 652
Db 618 AATATCTCTCTTTTCTTTTAAAGATTTTCAATGTCATGTCACAGAGAGGCCACCTTTT 677
Oy 653 gacattcatttcagagctgtgtgtgataaataagatcgacacatgcaagtgttcaat 712
Db 678 GACTGTGCTATTTCAGGTTGTGTGATGATGATAA-AAGATCGACCAATGGAGCTTTCTATG 726
Oy 713 gacttccaattggtccctgtatgttcaagatgtatcactcactccctgactgtactt 772
Db 737 ACCTTCAATTTGGCTCTGAGATGTCACAGATGTGTTGTTGTTGTTGTTGTTGTTG 796
Oy 773 caatgggagatgagaagtttccagagaactgaactgtggaataatgaccttcttact 832
Db 797 CAGTGGGAGATGAAATTTTCAAGAACTGAATCTGTG-AAATGACCTTTCTCTCAGCT 855
Oy 833 tgaagctactttaaatttgaaggtctgagacaaagaagaagatatacaagttgaagt 892
Db 856 TGAAGCTACTTTTAAATCTCGGGGTCTGACCAAAAGAAACATCAT--GTTTGTAGT 913
Oy 893 caagatgacagataagtgagatgaatgaactcactcaagaatgtgtcacttgaagaaa 952
Db 914 CAAGATGACATACAGAGAGAGACAGCTAATCTCCAAAGGTGCTTCACTGAGAGAA 973
Oy 953 ggcatttaagaatttttaaataatctgtcagaagaatcccgaaagaagttaatttcat 1012
Db 974 ACAAGTGTCTTGAGCAGACAGATTTCTCAGAGATGCCAGGAAGTTTAAATGTTAT 1033
Oy 1013 tagca---attaataaagctataacagagaatgaatgaacagagaactgtcctttt 1069
Db 1034 CAGCAGTATTAATAAGTTACTCATACAGAGTGTACGCAACAA-----CACTGCTTT 1088
Oy 1070 agatttatttacttttggccctggagataggttttaaatgacattgtctgacca 1129
Db 1089 TCATCTCTGTTTACTTTTGGCTGGGACATGGGCTTTTCAAGGACATCGTCTTACCA 1148
Oy 1130 gcttcaataaataaacaatttgcataa 1159
|||||

```

```

Db 1149 GCTTCAATTAATAAATATTTGTAATAA 1178

RESULT 13
AL532393/c 943 bp mRNA EST 13-FEB-2001
LOCUS AL532393 LTL_NFL001_NBC4 Homo sapiens cDNA clone CS0DM014YH24 3
DEFINITION prime, mRNA sequence.
ACCESSION AL532393
VERSION AL532393.1 GI:12795886
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 943)
AUTHORS Li, W.-B., Gruber, C., Jassie, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqrel@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..943
location/Qualifiers
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="CS0DM014YH24"
    /clone_id="LTL_NFL001_NBC4"
    /sex="male"
    /tissue_type="neuroblastoma cells"
    /lab_host="DH10B"
    /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dt) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by life technologies. Contact : Feng Liang life
Technologies, a division of invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@life.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 199 a 185 c 187 g 294 t 78 others
ORIGIN

Query Match 50.3%; Score 737; DB 106; Length 943;
Best Local Similarity 87.0%; Pred No. 1.7e-145;
Matches 824; Conservative 6; Mismatches 111; Indels 6; Gaps 6;

Oy 184 gagatcactcaaatgaaatgaacacacatcgaagaactcaagaatcaaatcactcgt 243
Db 943 GAGATTCACATCAAAAGGAAATGACACACATCTCAAGAACTCAAAAGATCATCTGT 884
Oy 244 caaagacagggtgttccaatgaatcactcagtttctcttgagggtcagaagattgt 303
Db 883 CAAAGACAGGGGTCTCCCAATGAATMCATCAGGTGCTCTTTGAGGTCAGAAATTTGCT 824
Oy 304 gataatcattctc-aaagaactgggaatggagaagaatgattgaattgaatca 362
Db 823 GATATATATCTCCAAAAGAACTGGAATGAGAGAAAGATGTGATTCAGAAATTTATCA 764
Oy 363 ggaacaacagggtgtcattcaacagtttagatattcttatttcttcttccct 422
Db 763 GGAGCAACAC-GGGGGKATTCACACAGTTAGATATCTTTTATTTTCTTTTCCCT 705
Oy 423 caatccttttattttaaataagttcttcttgtaatgagtggttcaaaagaaatga 482
Db 704 CAATCTTTTATTTTAAATAATAGTTC-TTGTGAANNGGNNNNCAAAACGAANNNGA 646
Oy 483 aaatggcaccctccttgaacacatcgttaattgaattcattgctcattatca 542
Db 645 AAGCGCACCCCAACNCNCNGAACAACANCGNNAANNNGNANNCNAGNCNANNA 586

```


QY 835 aagctacttttaaaatttgaggtctgaccacaaagaagaatcatcaggttgaaatca 894
 DB 842 AAGAACTTTTAAAGATTTGGGGGTGTGAGCAAGAAAGGGGGAATTCGGGGGGAAGGCC 901
 QY 895 agatgacgataaagtgatgagtaagtaagtaagtaagtaagtaagtaagtaagtaag 954
 DB 902 AGAAGATTAAGAAAGGGGTGGAGTAATTAAGTAATTAACCAAGAGGGGTGATCGAAGCAGG 961
 QY 955 catttaagaatttttaaa 972
 DB 962 GATGTAGAAATTTTATAA 979
 RESULT 15
 LOCUS BG540101 804 bp mRNA EST 03-APR-2001
 DEFINITION 602568937F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4693352 5',
 mRNA sequence.
 ACCESSION BG540101 GI:13532334
 VERSION BG540101.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 804)
 NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LICM1515 row: 1 column: 09
 High quality sequence stop: 785.
 Location/Qualifiers
 1..804
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4693352"
 /clone_id="NIH_MGC_77"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1:
 SfiI (ggcgccctggcc); Site: 2: SfiI (ggcgccctggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCGCCATTATGCCC-3, and 3' adaptor
 sequence: 5'-ATCTAGAGCGCGGCGCGCGCATG-3 (30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.9
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."
 BASE COUNT 227 a 150 c 184 g 243 t
 ORIGIN
 Query Match 50.0%; Score 732; DB 155; Length 804;
 Best Local Similarity 98.3%; Pred. No. 1.9e-144;
 Matches 771; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
 QY 2 gggaaagattttaaaccggagcagaggttctgttaccacggagcgcgtgtgagcga 61
 DB 17 GAGAAAGATTGTAAACCCGGAGGAGGTCTGCTTACCGAGCGCGTGTGCGGA 76
 QY 62 gaccgccgggtgaagccacgctcatcatgtctgacacgagaggaacacttaactgagg 121
 DB 77 GACCCCGGGGTGAAGCCACGTCATCATGTCTGACACAGAGGCAAAACCTTCACTGAGG 136

QY 122 acttgaggatgaagaagcagagtgatataatataaactcaagcttcgtgacgagatga 181
 DB 137 ACTTGGGGGATTAAGAAAGAGATGTAATATTAACTCAAAAGTCATTGGACAGATAGA 196
 QY 182 gtgagcttcactcaaaagtgaataatgacacacatctcaagaactcaagaatcatact 241
 DB 197 GTGAGATTCACTTCAAAAGTGAATAATGACAAACATCTCAAGAAACTCAAAAGATCACT 256
 QY 242 gtcaaaagacaggggtgttccaatgaattcaactcaggtttctcttgagggtcaggaattg 301
 DB 257 GTCAAAGACAGGGGTGTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 316
 QY 302 ctgataatcaactccaagaacttggaatgaggaagaagaatgagatgagatgagatgag 361
 DB 317 CTGATTAATCACTACTCAAAAGAACTGGGAATGAGAGAGATGATGATGATGATGATGAT 376
 QY 362 aggaacaaacgggggtgcatcacaagtttagatattcttatttatttcttcttcc 421
 DB 377 AGGAAACAAAGGGGGGTGATCAAGATTAGATATTCTTTTATTTTCTTTCTTTCTTTCC 436
 QY 422 tcaatccttttatttattttaaataatgtcttcttgtaagtgtgttcaaaacggatg 481
 DB 437 TCAATCCTTTTATTTTATTTTAAAAATAGTCTTTTGTAAATGTGTGTCAAAAGGATTTG 496
 QY 482 aaactgacaccccatctcttgaaacatctgtaattgaattcagtcagtcattatc 541
 DB 497 AAACGTGGACCCCATCTCTTTGAACATCTGTAATTGAATTCAGTGTCTATATTTC 556
 QY 542 atiatgtttgtttcaattgtgtgatttttgatgcaagccctcagtccttcaatt 601
 DB 557 ATATGTTGTGTTTCAATGTGTGATCTTGGTGAATCAAGCTCAGTCCCTTCATATT 616
 QY 602 accctccctttttaaataatgaatgaatgagcagaagaggtacacttttgaagagatgca 661
 DB 617 ACCCTCTCTTTTAAAAATTAAGTGTGACAGAGAGGTACCTTTTTCAGCATTTGCA 676
 QY 662 tttaagcctgtgtgtataataagaatgacacatgcaatgcaatgcaatgcaatgcaatgca 720
 DB 677 TTTTCAAGCTTGTGTGTATTAATTAAGATGACCAATGCAAGTGTCAATATGACTTTCCA 736
 QY 721 aattgcccctgattgtc-agcatgtgattacttcaactcctgagct-gtgaattcagttg 778
 DB 737 AATGGCCCTGATGTCTAGCATGTGATTAATCAATCTGAGTGTGATCTTTCAGTGG 796
 QY 779 gaga 782
 DB 797 GAGA 800

Search completed: August 15, 2001, 16:12:28
 Job time: 3239 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 16:36:04 ; Search time 72.32 seconds

(without alignments)
28.736 Million cell updates/sec

Title: US-09-484-964-2

Sequence: 1 MSQGEAKPRFEDGDKKQGF.....MEEDYIEYQEQTGSHSTV 101

Scoring table: BIOSUM62

Gap 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCYUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	223.5	42.6	104	2	US-08-853-974-3
2	223.5	42.6	104	4	US-09-172-988-3
3	217	41.3	126	2	US-08-853-974-1
4	217	41.3	126	4	US-09-172-988-1
5	101	19.2	412	2	US-08-755-584-2
6	101	19.2	412	3	US-09-192-611-2
7	76	14.5	352	3	US-08-854-764-2
8	76	14.5	352	5	PCT-US95-09377-2
9	73	13.9	75	1	US-08-350-884-35
10	73	13.9	75	1	US-08-709-173-35
11	73	13.9	75	2	US-08-709-177-35
12	73	13.9	75	2	US-08-833-678A-3
13	73	13.9	75	4	US-08-529-169A-3
14	73	13.9	76	1	US-08-232-815-2
15	73	13.9	76	1	US-08-350-906-2
16	73	13.9	76	2	US-09-070-060-8
17	73	13.9	76	3	US-09-357-746-8
18	73	13.9	76	5	PCT-US95-04536-2
19	73	13.9	78	2	US-08-505-486-94
20	73	13.9	78	3	US-08-801-028-94
21	73	13.9	78	3	US-09-340-154-94
22	73	13.9	78	5	PCT-US95-09338-94
23	73	13.9	78	5	PCT-US95-09339-94
24	73	13.9	103	2	US-08-771-201-9
25	73	13.9	114	2	US-08-771-201-10
26	73	13.9	128	2	US-08-557-309B-41
27	73	13.9	128	3	US-08-834-306-41

28	73	13.9	128	4	US-08-993-674A-41	Sequence 41, Appl
29	73	13.9	147	2	US-08-771-201-11	Sequence 11, Appl
30	73	13.9	156	2	US-09-070-060-7	Sequence 7, Appl
31	73	13.9	156	3	US-09-051-969A-3	Sequence 3, Appl
32	73	13.9	156	3	US-09-051-969A-4	Sequence 4, Appl
33	73	13.9	156	3	US-09-357-746-7	Sequence 7, Appl
34	73	13.9	229	2	US-08-726-306A-23	Sequence 23, Appl
35	73	13.9	229	3	US-08-840-146-20	Sequence 20, Appl
36	73	13.9	229	3	US-09-360-220-20	Sequence 20, Appl
37	73	13.9	533	6	5510474-2	Patent No. 5510474
38	73	13.9	1121	1	US-07-789-915A-2	Sequence 2, Appl
39	73	13.9	1121	1	US-08-005-002C-2	Sequence 2, Appl
40	73	13.9	1121	1	US-08-487-203A-2	Sequence 2, Appl
41	71	13.5	76	1	US-08-450-834-6	Sequence 6, Appl
42	71	13.5	77	6	5510474-3	Patent No. 5510474
43	71	13.5	533	1	US-08-462-092-2	Sequence 2, Appl
44	71	13.5	533	3	US-08-746-822-2	Sequence 2, Appl
45	71	13.5	533	3	US-09-094-350-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-853-974-3
: Sequence 3, Application US/08853974
: Patent No. 5840534
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Shah, Purvi
: TITLE OF INVENTION: HUMAN SWT3-LIKE PROTEIN
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTED for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/853,974
: FILING DATE: Filed Herewith
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0289 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 104 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: Genbank
: CLONE: 1707372
: US-08-853-974-3

Query Match 42.6%; Score 223.5; DB 2; Length 104;
Best Local Similarity 51.2%; Pred. No. 4,2e-19;
Matches 44; Conservative 14; Mismatches 25; Indels 3; Gaps 1;


```

QY      6 AKPSTEDIG-----DRKQGEYIKAVIGDSE--IHPKKVTKTKTKTE 49
          ||| |||
Db      310 ATPSTLRKGVADIIDCVLIASSSEATFESQGLRRLRVQGEKHNQLLEIISDPSLTKLMS 369
          ||| |||
QY      50 SYCCRGVPMNSLFLFEQGRIALDNHPKELGMEEDVEIY 91
          ||| |||
Db      370 HTEAMGSLGKLSFFEDGKLPADGLGSSGDIIEW 411
          ||| |||

```

```

RESULT 7
US-08-854-764-2
: Sequence 2, Application US/08854764
: Patent No. 6103500
: GENERAL INFORMATION:
: APPLICANT: Innes, Michael
: APPLICANT: Creasey, Abia
: TITLE OF INVENTION: Production of Tissue Factor Pathway
: TITLE OF INVENTION: Inhibitor
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton St.
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/854,764
: FILING DATE: 12-May-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/286,530
: FILING DATE: 05-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Saveriede, Paul B.
: REGISTRATION NUMBER: 36,914
: REFERENCE/DOCKET NUMBER: 0991.001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 510-601-2585
: TELEFAX: 510-655-3542
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 352 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-854-764-2

```

```

Query Match          14.5% ; Score 76 ; DB 3 ; Length 352 ;
Best Local Similarity 21.2% ; Pred. No. 0.39 ;
Matches 14 ; Conservative 19 ; Mismatches 33 ; Indels 0 ; Gaps 0 ;

QY 34 IHFYKMTHTLKLKESYCCORQGVPMNSLRLFLFSQRIADNHTPKELMEEDVEIVYQOE 93
    | . . . . . | . . . . . | . . . . . | . . . . . |
Db 13 IITLEESDRTIDNVKSKIQDKEGIPDDQRLIFAGKOLEDSRTSLSDYNIQKESLHLVLR 72

QY 94 OTGGHS 99
    | | |
Db 73 LRGGDS 78

RESULT 8
PCT-US95-09377-2
; Sequence 2, Application PC/TUS9509377
; GENERAL INFORMATION:
; APPLICANT: CHIRON CORPORATION

```

? TITLE OF INVENTION: Production of Tissue Factor Pathway
 ? TITLE OF INVENTION: Inhibitor
 ? NUMBER OF SEQUENCES: 7
 ? CORRESPONDENCE ADDRESSES:
 ? ADDRESSEE: Chiron Corporation
 ? STREET: 4560 Horton St.
 ? CITY: Emeryville
 ? STATE: CA
 ? COUNTRY: USA
 ? ZIP: 94608
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patentin Release #1.0, Version #1.30B
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: PCT/US95/09377
 ? FILING DATE: 25-JULY-1995
 ? CLASSIFICATION:
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Savereide, Paul B.
 ? REGISTRATION NUMBER: 36,914
 ? REFERENCE/DOCKET NUMBER: 0991.100
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 510-601-2585
 ? TELEFAX: 510-653-3542
 ? INFORMATION FOR SEQ ID NO: 2:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 352 amino acids
 ? TYPE: amino acid
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: protein

	Query Match	14.5%	Score 76;	DB 5;	Length 352;
	Best Local Similarity	21.2%;	Pred. No.	0.39;	
	Matches 14;	Conservative 19;	Mismatches 33;	Indels 0;	Gaps 0;
OY	34	IHFVKMTTHLKLKESYCORGVPMNSLRLEFGORIADNHTPRELCMEDEEYVQOE	93		
	: :	: :	: :	: :	:
DG	13	ITLEESSDTIDNVKSKIQDKEGIPDQQRILIFAKGLEDPGRFLSDVIYTKRESTLHVLR	72		
OY	94	QTGHS	99		
DG	73	LRCGDS	78		

RESULT 9
US-08-350-884-35
Sequence 35, Application US/08350884
Patent No. 5585258
GENERAL INFORMATION:
APPLICANT: HOUGHTON, MICHAEL
APPLICANT: CHOO, OUI LIM
APPLICANT: KUO, GEORGE
TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/350,884
FILING DATE: 06-DEC-1994

Matches	137	Conservative	197	Mismatches	327	Indels	07	Gaps	07
34	IIHKVAMTHLKKLKESTCQKRGVPMNSLRFLFESQKADNHPRKELGMEEDVIEVYQ	93							
QY	IIHKVAMTHLKKLKESTCQKRGVPMNSLRFLFESQKADNHPRKELGMEEDVIEVYQ	93							
12	ITLVESSTPTINNVSKTDGKRGTPPDQDRIIPAGKQIIPGQRTISPVNQTQKSTNIHLVR	71							
b	ITLVESSTPTINNVSKTDGKRGTPPDQDRIIPAGKQIIPGQRTISPVNQTQKSTNIHLVR	71							

OY 94 QTGG 97
||
Db 72 LRGG 75

RESULT 12

US-08-833-678A-3
; Sequence 3, Application US/08833678A
; Patent No. 5989905

GENERAL INFORMATION:

APPLICANT: HOUGHTON, MICHAEL
APPLICANT: CHOO, OUI-LIM

APPLICANT: HAN, JANG

APPLICANT: CHOE, JOONHO
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION
STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville
STATE: California

COUNTRY: U.S.A.
ZIP: 94662-8097

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/833,678A
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/529,169
FILING DATE: 15-SEP-1995

ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0100.005

TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-3274

TELEFAX: (510) 655-3542
TELEX: n/a

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 75 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-833-678A-3

Query Match 13.9%; Score 73; DB 2; Length 75;
Best Local Similarity 20.3%; Pred. No. 0.12;
Matches 13; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

OY 34 IHFKYMTHTLKKESYQROGVPMNSLRFEEGRIADNHTPKELGMEEDVEVYOE 93
||
Db 12 ILEVESSDTIYNVSKIDKGIIPDQORLIFAGKQLEDGRTLSYNIQKESSTLHLVLR 71

OY 94 QTGG 97
||
Db 72 LRGG 75

RESULT 13

US-08-529-169A-3
; Sequence 3, Application US/08529169A
; Patent No. 6194140

GENERAL INFORMATION:

APPLICANT: HOUGHTON, MICHAEL

APPLICANT: CHOO, OUI-LIM
APPLICANT: HAN, JANG

APPLICANT: CHOE, JOONHO
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION
STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville
STATE: California

COUNTRY: U.S.A.
ZIP: 94662-8097

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/529,169A
FILING DATE: 15-SEP-1995

CLASSIFICATION: 4325
ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0100.005
TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 923-3274
TELEFAX: (510) 655-3542

TELEX: n/a
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 75 amino acids

TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-529-169A-3

Query Match 13.9%; Score 73; DB 4; Length 75;
Best Local Similarity 20.3%; Pred. No. 0.12;
Matches 13; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

OY 34 IHFKYMTHTLKKESYQROGVPMNSLRFEEGRIADNHTPKELGMEEDVEVYOE 93
||
Db 12 ILEVESSDTIYNVSKIDKGIIPDQORLIFAGKQLEDGRTLSYNIQKESSTLHLVLR 71

OY 94 QTGG 97
||
Db 72 LRGG 75

RESULT 14

US-08-232-815-2
; Sequence 2, Application US/08232815
; Patent No. 5503977

GENERAL INFORMATION:

APPLICANT: Johnson, Nils
APPLICANT: Varshavsky, Alexander

TITLE OF INVENTION: SPLIT UBIQUITIN PROTEIN SENSOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSEE: Kevin M. Farrell
STREET: P.O. Box 999

CITY: York Harbor
STATE: Maine

COUNTRY: USA
ZIP: 03911

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

Thu Aug 16 08:48:53 2001

us-09-484-964-2.rai

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 16:16:34 ; Search time 63.2 Seconds
(without alignments)
96.883 Million cell updates/sec

Title: US-09-484-964-2

Sequence: 1 MSDEAKPSTEDLDKKKQGE.....MEEDVLEVYQEQTGHSTV 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	525	100.0	101	19	AAW60079
2	522	99.4	101	20	AAW60079
3	522	99.4	101	21	AAW60079
4	522	99.4	102	19	AAW74776
5	239.5	45.6	95	20	AAW87985
6	232	44.2	95	20	AAW87985
7	223.5	42.6	100	21	AAW87985
8	221	42.1	103	21	AAW87985
9	221	42.1	103	21	AAW87985
10	221	42.1	133	21	AAW87985
11	218.5	41.6	100	20	AAW87987

12	218	41.5	99	21	AAW60079	zebra mays protein f
13	217	41.3	126	20	AAW60079	Human SMT3-like pr
14	175.5	33.4	111	21	AAW60079	Arabidopsis thalia
15	175.5	33.4	145	21	AAW60079	Human prostate can
16	147	28.0	83	21	AAW60079	Arabidopsis thalia
17	147	28.0	83	21	AAW60079	Arabidopsis thalia
18	147	28.0	108	21	AAW60079	Arabidopsis thalia
19	147	28.0	108	21	AAW60079	Arabidopsis thalia
20	147	28.0	142	21	AAW60079	Arabidopsis thalia
21	147	28.0	143	21	AAW60079	Arabidopsis thalia
22	133	25.3	57	21	AAW60079	Arabidopsis thalia
23	127	24.2	61	21	AAW60079	Arabidopsis thalia
24	111	21.1	66	21	AAW60079	Arabidopsis thalia
25	104	19.8	122	21	AAW60079	Arabidopsis thalia
26	104	19.8	215	21	AAW60079	Arabidopsis thalia
27	104	19.8	244	21	AAW60079	Arabidopsis thalia
28	101	19.2	153	21	AAW60079	zebra mays protein f
29	101	19.2	190	21	AAW60079	zebra mays protein f
30	101	19.2	412	18	AAW60079	Mouse NF-AT intera
31	101	19.2	412	20	AAW60079	Human NIP45 protel
32	100	19.0	419	20	AAW60079	Arabidopsis thalia
33	98	18.7	154	21	AAW60079	Arabidopsis thalia
34	98	18.7	156	21	AAW60079	Arabidopsis thalia
35	98	18.7	169	21	AAW60079	Arabidopsis thalia
36	96	18.3	78	21	AAW60079	Arabidopsis thalia
37	94.5	18.0	153	22	AAW60079	Amino acid sequenc
38	94.5	18.0	183	22	AAW60079	Amino acid sequenc
39	94.5	18.0	220	22	AAW60079	Amino acid sequenc
40	94.5	18.0	250	22	AAW60079	Amino acid sequenc
41	94.5	18.0	262	22	AAW60079	Amino acid sequenc
42	94.5	18.0	296	22	AAW60079	Amino acid sequenc
43	94.5	18.0	451	22	AAW60079	Amino acid sequenc
44	94.5	18.0	464	22	AAW60079	Amino acid sequenc
45	94.5	18.0	464	22	AAW60079	Amino acid sequenc

ALIGNMENTS

RESULT	1	AAW60079	standard; Protein; 101 AA.
ID	AAW60079	standard; Protein; 101 AA.	
AC	AAW60079		
DT	11-SEP-1998	(first entry)	
DE	Homo sapiens sentrin-1 polypeptide.		
KW	sentrin-1; protection; tumour necrosis factor; tnf; apoptosis;		
KW	Fas/Apo-induced; tumour cell death; induction; tumour aggressiveness;		
KW	detection; determination.		
OS	Homo sapiens.		
PN	WO9820038-A1.		
PD	14-MAY-1998.		
PF	05-NOV-1997;	97WO-US20344.	
PR	05-NOV-1996;	96US-0030302.	
TX	(TEXAS) UNIV TEXAS SYSTEM.		
PI	Yeh ETH;		
DR	WPI: 1998-286668/25.		
XX	N-PSDB; AAW34564.		
XX	New isolated sentrin polypeptide(s) - which inhibit TNF receptor or		
PT	Fas/Apo-induced apoptosis, used to develop products for inducing		
PT	cell death in tumours		

XX Claim 4; Page 78; 120pp; English.

PS The sequence is that of the sentrin-1 polypeptide.
 CC Sentrin polypeptides have the ability to protect or guard
 CC cells from tumour necrosis factor (TNF) or Fas/Apo induced
 CC cell death (apoptosis). Inhibitors of the sentrin polypeptides,
 CC e.g. antibodies, can be used for inducing cell death,
 CC particularly in tumours. The products can also be used for
 CC determining the aggressiveness of a tumour and for detection and
 CC isolation of products. The sentrin polypeptide can also be used to
 CC detect a ubiquitin conjugating enzyme polypeptide or PML polypeptide.
 XX Sequence 101 AA;

Query Match 100.0%; Score 525; DB 19; Length 101;
 Best Local Similarity 100.0%; Pred. No. 1.2e-54;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGODSSEIHFKVMTTHLKKLKESYCORGVPMN 60
 DB 1 msdgaekpstedldgdkkegeyiklvigdsseihfkvmtthlkkkkesyqgrgvpmn 60
 QY 61 SIRFLFEGORLADNHTPKELGMEEDVIEVYQGTGSHSTV 101
 DB 61 sirflfegqrladnhtpkelgmeedvievyqgtgshstv 101

RESULT 2

AAW87984
 ID AAW87984 standard; peptide; 101 AA.

XX AAW87984;
 AC
 XX 29-APR-1999 (first entry)

DE Ubiquitin-like domain of the protein SUMO1.

KW Ubiquitin; ubiquitin/proteasome pathway; degradation signal; UBL domain;
 KM intracellular protein degradation; proteasome; ubiquitin-like domain;
 KM fusion protein; drug resistance; malignant cell; purification;
 KM SUMO1.

OS Mammalia.

PN WO9857978-A1

PD 23-DEC-1998.

PF 19-JUN-1998; 98WO-US12846.

PR 19-JUN-1997; 97US-0050171.

PA (UYNE-) UNIV NEW JERSEY.

PI Madura K;

WPI; 1999-095322/08.

PT Ubiquitin-like domains as fusion or immobilised proteins - useful,
 PT respectively, for assessing the proliferative potential of malignant
 PT cells or to purify proteasome complexes

PS Claim 2; Page 25; 84pp; English.

XX The present sequence represents a ubiquitin-like domain of a
 CC mammalian protein. The ubiquitin/proteasome pathway is a major
 CC pathway for the intracellular degradation of proteins, where the
 CC ubiquitin acts as a degradation signal. Proteins which bind the
 CC proteasome (a multisubunit complex) have ubiquitin-like (UBL)
 CC domains. When the UBL domains are fused to a reporter protein,
 CC they can be used to assess the proliferative potential and select

CC drug resistance of malignant cells. The UBL domain also allows the
 CC purification of proteasomes for the isolation and characterisation
 CC of novel subunits of the proteasome.

SQ Sequence 101 AA;

Query Match 99.4%; Score 522; DB 20; Length 101;
 Best Local Similarity 99.0%; Pred. No. 2.6e-54;
 Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGODSSEIHFKVMTTHLKKLKESYCORGVPMN 60
 DB 1 msdgaekpstedldgdkkegeyiklvigdsseihfkvmtthlkkkkesyqgrgvpmn 60
 QY 61 SIRFLFEGORLADNHTPKELGMEEDVIEVYQGTGSHSTV 101
 DB 61 sirflfegqrladnhtpkelgmeedvievyqgtgshstv 101

RESULT 3

AAV49967
 ID AAV49967 standard; Protein; 101 AA.

XX AAV49967;

AC
 XX 03-FEB-2000 (first entry)

DE Human sentrin protein sequence.

KW Human; sentrin; antisense oligonucleotide; phosphorothioate;
 KM inhibition; modulation; expression; diagnosis.

OS Homo sapiens.

PN US5985664-A.

PD 16-NOV-1999.

PF 17-DEC-1998; 98US-0213768.

PR 17-DEC-1998; 98US-0213768.

PA (ISIS-) ISIS PHARM INC.

PI Baker BF, Cowsett LM;

WPI; 2000-022284/02.

DR N-PSDB; AA235859.

PT Antisense compound which modulates human sentrin expression, useful for
 PT treating diseases associated with sentrin expression -
 PS Example 13; Column 39-42; 29pp; English.

XX The present invention describes an antisense compound (I) 8-30
 CC nucleotides long targeted to a nucleic acid molecule encoding human
 CC sentrin. The antisense compound comprises a phosphorothioate antisense
 CC oligonucleotide which inhibits expression of human sentrin. (I) is
 CC useful for inhibiting expression of sentrin in human cells or tissues
 CC in vitro, for treating humans or other animals suspected of having or
 CC being prone to a disease associated with sentrin expression. (I) can
 CC also be used for research or diagnostic purposes. The present
 CC sequence represents human sentrin.

SQ Sequence 101 AA;

Query Match 99.4%; Score 522; DB 21; Length 101;
 Best Local Similarity 99.0%; Pred. No. 2.6e-54;
 Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGODSSEIHFKVMTTHLKKLKESYCORGVPMN 60

```
Db 1 mdsqekpstedigdkkeeyikivlgdselrkvmcthlkklksyccqrgvpm 60
OY 61 SLRFLFEGORIANHTPKELGMEEDVLEVYQEQTGHSTV 101
Db 61 SLRFLFEGORIANHTPKELGMEEDVLEVYQEQTGHSTV 101

RESULT 4
AAW74776
ID AAW74776 standard; Protein; 102 AA.
AC AAW74776;
XX 25-JAN-1999 (first entry)
DT Human secreted protein encoded by gene 47 clone HOCAY75.
XX
DE Human; secreted protein; testis; tumour; foetal brain tissue;
KW fusion protein; cancer; central nervous system; seizure;
KM diagnosis; neurodegenerative disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 102
XX label= unknown
XX
PN MO9839448-A2.
PD 11-SEP-1998.
XX
XX 06-MAR-1998; 98MO-US04493.
PF
XX 02-OCT-1997; 97US-0061060.
PR 07-MAR-1997; 97US-0038621.
PR 07-MAR-1997; 97US-0040161.
PR 07-MAR-1997; 97US-0040162.
PR 07-MAR-1997; 97US-0040163.
PR 07-MAR-1997; 97US-0040333.
PR 07-MAR-1997; 97US-0040334.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043356.
PR 11-APR-1997; 97US-0043359.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.

PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056639.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056846.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057659.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.

(PHUMAN) HUMAN GENOME SCI INC.
XX
XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ehner R, Endress GA;
PI Feng P, Petrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Ryaw H, Latleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
DR MPI: 1998-506364/43.
DR N-PDB: AAV59557.
XX
PT New isolated human genes and the secreted polypeptide(s) they encode
- useful for diagnosis and treatment of e.g. cancers, neurological
```

PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 566-567; 721pp; English.

XX This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 47 from the human cDNA clone HOGAV75
CC (deposited as clone ATCC 97899 and ATCC 209045).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: AAV59511-595812; amino acid sequences AAW74731-W75026)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAV59511 for described uses).

XX Sequence 102 AA;

Query Match 99.4%; Score 522; DB 19; Length 102;
Best Local Similarity 99.0%; Pred. No. 2.7e-54;
Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGDSSSEIFKFKMTTHLKKLESYQGRGVPMN 60
DB 1 msdgaekpstedldgkkggeyiklvigdsesselfkfkmtthlkklesycqrgvpmn 60
OY 61 SLRFLFEGQRIADNHTPKRELGMEEDEVIEVYQEOGTGHSTV 101
DB 61 slrflfegqriadnhtpkrelgmeeedvievyyqegtghstv 101

RESULT 5

AAW87986 standard; peptide: 95 AA.

XX AAW87986;

DT 29-APR-1999 (first entry)

DE Ubiquitin-like domain of the yeast protein SMT3A.

KW Ubiquitin; ubiquitin/proteasome pathway; degradation signal; Ubl domain;
KM intracellular protein degradation; proteasome; ubiquitin-like domain;
KW fusion protein; drug resistance; malignant cell; purification;
KM yeast; SMT3A.

OS Saccharomyces sp.

XX MO9857978-A1.

PD 23-DEC-1998.

PF 19-JUN-1998; 98WO-US12846.

PR 19-JUN-1997; 97US-0050171.

PA (UYNE-) UNIV NEW JERSEY.

PI Madura K;

XX WPI; 1999-095322/08.

PT Ubiquitin-like domains as fusion or immobilised proteins - useful,
PT respectively, for assessing the proliferative potential of malignant
PT cells or to purify proteasome complexes
XX
PS Claim 2; Page 25; 84pp; English.

CC The present sequence represents a ubiquitin-like domain of the
CC yeast protein SMT3A. The ubiquitin/proteasome pathway is a major
CC pathway for the intracellular degradation of proteins, where the
CC ubiquitin acts as a degradation signal. Proteins which bind the
CC proteasome (a multisubunit complex) have ubiquitin-like (Ubl)
CC domains. When the Ubl domains are fused to a reporter protein,
CC they can be used to assess the proliferative potential and select
CC drug resistance of malignant cells. The Ubl domain also allows the
CC purification of proteasomes for the isolation and characterisation
CC of novel subunits of the proteasome.

XX Sequence 95 AA;

Query Match 45.6%; Score 239.5; DB 20; Length 95;
Best Local Similarity 47.4%; Pred. No. 6.6e-21;
Matches 46; Conservative 20; Mismatches 26; Indels 5; Gaps 1;

OY 1 MSDOAKPSTEDLDGKKGGEYIKLVIGDSSSEIFKFKMTTHLKKLESYQGRGVPMN 60
DB 1 mseeakpke-----gkktendhnlkvaqdgsvqfklrhtslsklmkayceerglsmr 55
OY 61 SLRFLFEGQRIADNHTPKRELGMEEDEVIEVYQEOGTGHG 97
DB 56 qirfrfdgqplnetdtpaglmmededtdvdfqgtg 92

RESULT 6

AAW87985 standard; peptide: 95 AA.

XX AAW87985;

DT 29-APR-1999 (first entry)

DE Ubiquitin-like domain of the yeast protein SMT3B.

KW Ubiquitin; ubiquitin/proteasome pathway; degradation signal; Ubl domain;
KM intracellular protein degradation; proteasome; ubiquitin-like domain;
KW fusion protein; drug resistance; malignant cell; purification;
KM yeast; SMT3B.

OS Saccharomyces sp.

XX WO9857978-A1.

PD 23-DEC-1998.

PF 19-JUN-1998; 98WO-US12846.

PR 19-JUN-1997; 97US-0050171.

PA (UYNE-) UNIV NEW JERSEY.

PI Madura K;

XX WPI; 1999-095322/08.

PT Ubiquitin-like domains as fusion or immobilised proteins - useful,
PT respectively, for assessing the proliferative potential of malignant
PT cells or to purify proteasome complexes
XX
PS Claim 2; Page 25; 84pp; English.

CC The present sequence represents a ubiquitin-like domain of the
CC yeast protein SMT3B. The ubiquitin/proteasome pathway is a major
CC pathway for the intracellular degradation of proteins, where the
CC ubiquitin acts as a degradation signal. Proteins which bind the
CC proteasome (a multisubunit complex) have ubiquitin-like (Ubl)
CC domains. When the Ubl domains are fused to a reporter protein,
CC they can be used to assess the proliferative potential and select
CC drug resistance of malignant cells. The Ubl domain also allows the
CC purification of proteasomes for the isolation and characterisation

Query Match	Similarity	Score	DB	Length
Best Local	51.2%	223.5	DB 21	100
Matches 44	Conservative 14	Mismatches 25	Indels 3	Gaps 1
QY	15 DKRQGE---YIKRKIVGDSSEIHFKVMTTHUKLKESYGCORQGVPMNSLRFEGORI 71			
Db	8 dkrrpgggagahlnkvvgqgqnevfikrstqlklnmaycdtrsvdmstaflltdgrtl 67			
QY	72 ADNHPTKELGMEEDYEVYQEQTG 97			
Db	68 raeqtdelmedgedelamdlnbtg 93			
RESULT 8				
AA011141				
ID	AA011141 standard; Protein; 103 AA.			
XX				
AC	AA011141;			
XX				
DT	17-OCT-2000 (first entry)			
XX				
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 9747.			
XX				
KW	Protein identification; signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
XX	termination sequence.			
OS	Arabidopsis thaliana.			
PN	EP1033405-A2.			
XX				
PD	06-SEP-2000.			
XX				
PE	25-FEB-2000; 2000EP-0301439.			
XX				
PR	25-FEB-1999; 99US-0121825.			
PR	03-MAR-1999; 99US-0123180.			
PR	09-MAR-1999; 99US-0123548.			
PR	23-MAR-1999; 99US-0125788.			
PR	25-MAR-1999; 99US-0126264.			
PR	29-MAR-1999; 99US-0126785.			
PR	01-APR-1999; 99US-0127462.			
PR	06-APR-1999; 99US-0128234.			
PR	08-APR-1999; 99US-0128714.			
PR	16-APR-1999; 99US-0129845.			
PR	19-APR-1999; 99US-0130077.			
PR	21-APR-1999; 99US-0130449.			
PR	23-APR-1999; 99US-0130510.			
PR	23-APR-1999; 99US-0130891.			
PR	28-APR-1999; 99US-0131449.			
PR	30-APR-1999; 99US-0132048.			
PR	30-APR-1999; 99US-0132407.			
PR	04-MAY-1999; 99US-0132484.			
PR	05-MAY-1999; 99US-0132485.			
PR	06-MAY-1999; 99US-0132486.			
PR	06-MAY-1999; 99US-0132487.			
PR	07-MAY-1999; 99US-0132487.			
PR	11-MAY-1999; 99US-0132863.			
PR	14-MAY-1999; 99US-0134218.			
PR	14-MAY-1999; 99US-0134219.			
PR	14-MAY-1999; 99US-0134221.			
PR	14-MAY-1999; 99US-0134370.			
PR	14-MAY-1999; 99US-0134370.			
PR	18-MAY-1999; 99US-0134768.			
PR	19-MAY-1999; 99US-0134941.			
PR	20-MAY-1999; 99US-0135353.			
PR	21-MAY-1999; 99US-0135353.			
PR	24-MAY-1999; 99US-0135629.			
PR	25-MAY-1999; 99US-0136021.			
PR	27-MAY-1999; 99US-0136392.			
PR	28-MAY-1999; 99US-0136782.			
PR	01-JUN-1999; 99US-0137222.			

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145917.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151438.
PR 31-AUG-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 07-SEP-1999; 99US-0153070.
PR 10-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

42.1%; Score 221; DB 21; Length 103;

PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	15-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158222.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	18-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Oy		15	DK--CGEYIKLAKTIGDSSIEHKVMTTHLKTKJCSYCOPGVPMNSLRFLFEGGRIA	72
Dd		8	dKpdqgahlnlkxkgqgnevfirkrstqklkkmaycdrgsvdnslafldgrlr	67
Oy		73	DNHTPEKLGMEEEDVIEYYEQBTGS	97
Dd		68	aegtpdelemedgedelamlnhtgtg	92
RESULT		10		
ID	AAG46643			
	AAG46643 standard; Protein: 153 AA.			
XX				
AC	AAG46643;			
DT	18-OCT-2000 (first entry)			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 58703.			
KW	Protein identification assay; signal transduction pathway; metabolic pathway; hydridisation assay; genetic mapping; gene expression control; promoter termination sequence.			
KM				
XX	Arabidopsis thaliana.			
OS				
XX	EP1033405-A2.			
PN				
PD	06-SEP-2000.			
XX				
Pf	25-FEB-2000; 2000EP-0301439.			
XX				
PR	25-FEB-1999; 99US-0121825.			
PR	05-MAR-1999; 99US-0123180.			
PR	09-MAR-1999; 99US-0123548.			
PR	23-MAR-1999; 99US-0125788.			
PR	25-MAR-1999; 99US-0126264.			
PR	29-MAR-1999; 99US-0126785.			
PR	01-APR-1999; 99US-0127462.			
PR	06-APR-1999; 99US-0128234.			
PR	08-APR-1999; 99US-0128714.			
PR	16-APR-1999; 99US-0129845.			
PR	19-APR-1999; 99US-0130077.			
PR	21-APR-1999; 99US-0130449.			
PR	23-APR-1999; 99US-0130510.			
PR	23-APR-1999; 99US-0130891.			
PR	28-APR-1999; 99US-0131449.			
PR	30-APR-1999; 99US-0132048.			
PR	30-APR-1999; 99US-0132407.			
PR	04-MAY-1999; 99US-0132484.			
PR	05-MAY-1999; 99US-0132485.			
PR	06-MAY-1999; 99US-0132486.			
PR	06-MAY-1999; 99US-0132487.			
PR	07-MAY-1999; 99US-0132863.			
PR	11-MAY-1999; 99US-0134256.			
PR	14-MAY-1999; 99US-0134218.			
PR	14-MAY-1999; 99US-0134219.			
PR	14-MAY-1999; 99US-0134370.			
PR	14-MAY-1999; 99US-0134768.			
PR	18-MAY-1999; 99US-0134941.			
PR	19-MAY-1999; 99US-0135124.			
PR	20-MAY-1999; 99US-0135353.			
PR	21-MAY-1999; 99US-0135629.			
PR	24-MAY-1999; 99US-0136021.			
PR	25-MAY-1999; 99US-0136382.			
PR	27-MAY-1999; 99US-0136782.			
PR	28-MAY-1999; 99US-0137222.			
PR	01-JUN-1999; 99US-0137528.			
PR	03-JUN-1999; 99US-0137502.			
PR	04-JUN-1999; 99US-0137724.			
PR	07-JUN-1999; 99US-0138094.			
PR	08-JUN-1999; 99US-0138540.			
PR	10-JUN-1999; 99US-0138847.			
PR	10-JUN-1999; 99US-0138847.			

```
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139497.
PR 18-JUN-1999; 99US-0139497.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145313.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146386.
PR 03-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.1%; Score 221; DB 21; Length 153;
Best Local Similarity 51.8%; Pred. No. 1.9e-18;
Matches 44; Conservative 13; Mismatches 26; Indels 2; Gaps 1;
```

QY 73 DNHTEKELMEEDVEYEQOTGG 97
 ||||| : : : : :
 Db 118 aeqtpdelemgedeidamhqtg 142

RESULT 11

AAW87987
 ID AAW87987 standard; peptide: 100 AA.

AAW87987;
 AC

DT 29-APR-1999 (first entry)

XX Ubiquitin-like domain of the yeast protein SMT3.

XX Ubiquitin; ubiquitin/proteasome pathway; degradation signal; Ubl domain;

KW Intracellular protein degradation; proteasome; ubiquitin-like domain;

KW fusion protein; drug resistance; malignant cell; purification;

KW yeast; SMT3.

XX Saccharomyces sp.

PN MO9857978-A1.

XX 23-DEC-1998.

PF 19-JUN-1998; 98WO-US12846.

PR 19-JUN-1997; 97US-0050171.

XX (UYNE-) UNIV NEW JERSEY.

PI Madura K;

XX WPI; 1999-095322/08.

XX Ubiquitin-like domains as fusion or immobilised proteins - useful,

XX respectively, for assessing the proliferative potential of malignant

XX cells or to purify proteasome complexes

XX Claim 2; Page 25; 84pp; English.

XX The present sequence represents a ubiquitin-like domain of the

XX yeast protein SMT3. The ubiquitin/proteasome pathway is a major

XX pathway for the intracellular degradation of proteins, where the

XX ubiquitin acts as a degradation signal. Proteins which bind the

XX proteasome (a multisubunit complex) have ubiquitin-like (Ubl)

XX domains. When the Ubl domains are fused to a reporter protein,

XX they can be used to assess the proliferative potential and select

XX drug resistance of malignant cells. The Ubl domain also allows the

XX purification of proteasomes for the isolation and characterisation

XX of novel subunits of the proteasome.

XX Sequence 100 AA;

XX Query Match 41.6%; Score 218.5; DB 20; Length 100;

XX Best Local Similarity 48.5%; Pred. No. 2,2e-18;

XX Matches 49; Conservative 19; Mismatches 26; Indels 7; Gaps 3;

QY 1 MSD---OBAKPEEDLDKROGEYIKLVIGDSSEIHFKVMTTKLKKECYCOROG 56
 ||| ||||| : : : ||||| ||||| : : : ||||| : : : |||||
 Db 1 msdsevngeakpevkr--evkpeethlnlkv--sdgseifkktktprlrimeafakrg 57
 ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
 QY 57 VPMNSURKPLFEGQRIADNHTPRELMEEDVEYEQOTGG 97
 ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
 Db 58 kemdsirflydgiridqdpdelmedndlieahreq 98
 ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||

RESULT 12
 AAG41138
 ID AAG41138 standard; Protein: 99 AA.

XX AAG41138;

XX 18-OCT-2000 (first entry)

XX Zea mays protein fragment SEQ ID NO: 51144.

XX Protein identification: signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence; corn.

XX Zea mays subsp. mays.

PN EP103405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135253.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	24-JUN-1999;	99US-0140354.
PR	28-JUN-1999;	99US-0140695.
PR	29-JUN-1999;	99US-0140823.
PR	30-JUN-1999;	99US-0140991.
PR	01-JUL-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	02-JUL-1999;	99US-0142154.
PR	06-JUL-1999;	99US-0142055.
PR	08-JUL-1999;	99US-0142390.
PR	09-JUL-1999;	99US-0142803.
PR	12-JUL-1999;	99US-0142920.
PR	13-JUL-1999;	99US-0142977.
PR	14-JUL-1999;	99US-0143542.
PR	15-JUL-1999;	99US-0143624.
PR	16-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144633.
PR	21-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145119.
PR	22-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	03-AUG-1999;	99US-0146389.
PR	04-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	05-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	06-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	09-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	10-AUG-1999;	99US-0147935.
PR	11-AUG-1999;	99US-0148171.
PR	12-AUG-1999;	99US-0148319.
PR	13-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	16-AUG-1999;	99US-0148684.
PR	17-AUG-1999;	99US-0149368.
PR	18-AUG-1999;	99US-0149175.
PR	20-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149723.
PR	23-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151348.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157111.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158366.
PR	13-OCT-1999;	99US-0158299.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT	

PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137722.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0139119.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139452.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139452.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0150884.
PR	21-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151065.
PR	22-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151066.
PR	23-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151080.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151303.
PR	24-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151438.
PR	28-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0151930.
PR	29-JUN-1999;	99US-0140981.	PR	10-SEP-1999;	99US-0152363.
PR	30-JUN-1999;	99US-0141287.	PR	13-SEP-1999;	99US-0153070.
PR	01-JUL-1999;	99US-0141842.	PR	15-SEP-1999;	99US-0153758.
PR	02-JUL-1999;	99US-0142154.	PR	16-SEP-1999;	99US-0154018.
PR	06-JUL-1999;	99US-0142055.	PR	20-SEP-1999;	99US-0154039.
PR	09-JUL-1999;	99US-0142803.	PR	22-SEP-1999;	99US-0154779.
PR	09-JUL-1999;	99US-0142820.	PR	23-SEP-1999;	99US-0155139.
PR	12-JUL-1999;	99US-0142977.	PR	24-SEP-1999;	99US-0155486.
PR	14-JUL-1999;	99US-0143642.	PR	28-SEP-1999;	99US-0156458.
PR	15-JUL-1999;	99US-0143624.	PR	29-SEP-1999;	99US-0156596.
PR	16-JUL-1999;	99US-0144005.	PR	04-OCT-1999;	99US-0157117.
PR	19-JUL-1999;	99US-0144086.	PR	05-OCT-1999;	99US-0157759.
PR	19-JUL-1999;	99US-0144325.	PR	06-OCT-1999;	99US-0157865.
PR	19-JUL-1999;	99US-0144331.	PR	07-OCT-1999;	99US-0158029.
PR	19-JUL-1999;	99US-0144332.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144333.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144334.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144335.	PR	13-OCT-1999;	99US-0159294.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159325.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159329.
PR	21-JUL-1999;	99US-0144814.	PR	14-OCT-1999;	99US-0159330.
PR	21-JUL-1999;	99US-0145086.	PR	14-OCT-1999;	99US-0159637.
PR	21-JUL-1999;	99US-0145088.	PR	18-OCT-1999;	99US-0159638.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0159584.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160741.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160767.
PR	23-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160770.
PR	23-JUL-1999;	99US-0145218.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160814.
PR	26-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160980.
PR	27-JUL-1999;	99US-0145918.	PR	22-OCT-1999;	99US-0160981.
PR	27-JUL-1999;	99US-0145919.	PR	22-OCT-1999;	99US-0160981.
PR	28-JUL-1999;	99US-0145951.	PR	25-OCT-1999;	99US-0161404.
PR	02-AUG-1999;	99US-0146386.	PR	25-OCT-1999;	99US-0161405.
PR	02-AUG-1999;	99US-0146388.	PR	25-OCT-1999;	99US-0161406.
PR	03-AUG-1999;	99US-0147038.	PR	26-OCT-1999;	99US-0161359.
PR	04-AUG-1999;	99US-0147204.	PR	26-OCT-1999;	99US-0161360.
PR	04-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161361.
PR	05-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161992.
PR	05-AUG-1999;	99US-0147192.	PR	29-OCT-1999;	99US-0162142.

Query Match

33.4%; Score 175.5; DB 21; Length 111;

• • •

```
out_format : pfs
```

us-09-484-964-2.rge

OM of: US-09-15, 2001 5:57 PM

Date: Aug 1993
Results were produced by Computer

About: Res
Copyright (c)
Notes:

[illegible][illegible]

```
Query: length: 10
Query length: 1344157
Database: GenEmbl: *
Database sequences: 856060004
Database length: -856060004
Database time (sec): 1074.650000
Search time (sec):
```

	Documentacion	modifying prob
EScore Len	! u72722 Human gap mRNA for SMT3	
52	! X99586 H.sapiens ubiquitin-like F	
57	! X99586 Human ubiquitin-ubiquit	

[illegible][illegible]

us-09-484-964-2.1ge

[illegible]

FEATURES	Source
LOCUS	seq_documentation_block
DEFINITION	HS067122
ACCESSION	HS067122
KEYWORDS	Human
SOURCE	067122 ubiquitin-related
ORGANISM	067122.1 GI:1762972
REFERENCE	human.
AUTHORS	Euko Sapientis
TITLE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	Mahajan, R.K., Guan, T., Delphin, C., Gerace, L. and Melchior, F.
REFERENCE	to small ubiquitin-related protein SUMO-1-mRNA, complete cds.
AUTHORS	15-FEB-1997
TITLE	2 (access 1 to 816)
JOURNAL	Cell nuclear ubiquitin-related protein RanBP2 involved in targeting RanGAP1
REFERENCE	9714868 (1), 97-107 (1997)
AUTHORS	Mahajan, R.K., Guan, T., Delphin, C., Gerace, L. and Melchior, F.
TITLE	Submitted (16-AUG-1996)
JOURNAL	Research Institute, 10666 N
Source	LOCUS

BASE COUNT
ORIGIN

Length: 101
Gaps: 0

QVLEVYQEOQTGGHSTVⁿ

Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
US-09-484-964-2 x HS067122 ..

Align seg 1/1 to: HS067122 from: 1 to: 816

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysI 17
  |||
81 ATGCTCAGCAGGAGGCAAACTTCAACTGAGGAGGATTAAGAA 130
  |||
17 sGlnGlyGluTyrIleLeuLysValIleGlyGlnAspSerSerGluI 34
  |||
131 GGAAGGTGCAATATTAATCAAGTCATGGACAGATAGCAGTGA 180
  |||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
  |||
181 TTCACCTCAAGTGAATATGACACACATCTCAAGAACTCAAGAAATCA 230
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
  |||
231 TACTGTCAAGACAGGAGGTTCCAAATTCACCTGCTTCTTGA 280
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
  |||
281 GGGTCAGAGAAATGCTGATATCATCTCCAAAGAACTGGAGATGAGG 330
  |||
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyHisSerThr 100
  |||
331 AAGAAGATGTGATTCAAGTTATACGAAACAAAGGGGGGTCATTCAACA 380
  |||
101 Val 101
  |||
381 GTT 383

```

seq_name: gb_pr10:HS038784

seq_documentation_block:

LOCUS HS038784 1017 bp mRNA PRI 16-OCT-1996

DEFINITION Human ubiquitin-like protein mRNA, complete cds.

ACCESSION U08784

VERSION U08784.1 GI:1574947

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 1017)

Shen, Z., Pardington-Purtymun, P.E., Comeaux, J.C., Moyzis, R.K. and

Chen, D.J.

UBI1, a human ubiquitin-like protein associating with human

RAD51/RAD52 proteins

Genomics 36 (2), 271-279 (1996)

2 (bases 1 to 1017)

96411684

1.1017

Location/Qualifiers

source

1.1017

/organism="Homo sapiens"

/db_xref="taxon:9606"

/note="This cDNA was cloned with the yeast two hybrid

system. The protein interacts with human RAD51 and RAD52

proteins in yeast"

67.372

/note="Similar to ubiquitin and to yeast Sm13p (suppressor

of Mif2); Method: conceptual translation supplied by

author"

/codon_start=1

/protein_id="AAC50733.1"

/db_xref="GI:1574948"

BASE COUNT 310 a 180 c 213 g 314 t
ORIGIN
/translation="MSDQKAPSTEDIGDKKEGEYIKLVIGDSSSEIFKVKMTTHL
KKLESTCGRGPPMNSLRFLFPGQRIADNHTKELGMEEDVIEVYQGTGSHSTV"

alignment_scores:
Quality: 522.00 Length: 101
Ratio: 5.168 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
US-09-484-964-2 x HS038784 ..

Align seg 1/1 to: HS038784 from: 1 to: 1017

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysI 17
  |||
67 ATGCTCAGCAGGAGGCAAACTTCAACTGAGGAGGATTAAGAA 116
  |||
17 sGlnGlyGluTyrIleLeuLysValIleGlyGlnAspSerSerGluI 34
  |||
117 GGAAGGTGCAATATTAATCAAGTCATGGACAGATAGCAGTGA 166
  |||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
  |||
167 TTCACCTCAAGTGAATATGACACACATCTCAAGAACTCAAGAAATCA 216
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
  |||
217 TACTGTCAAGACAGGAGGTTCCAAATTCACCTGCTTCTTGA 266
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
  |||
267 GGGTCAGAGAAATGCTGATATCATCTCCAAAGAACTGGAGATGAGG 316
  |||
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyHisSerThr 100
  |||
317 AAGAAGATGTGATTCAAGTTATACGAAACAAAGGGGGGTCATTCAACA 366
  |||
101 Val 101
  |||
367 GTT 369

```

seq_name: gb_r01:AF033353

seq_documentation_block:

LOCUS AF033353 1187 bp mRNA ROD 06-FEB-1998

DEFINITION Mus musculus ubiquitin-homology domain protein (Ubi1) mRNA,

complete cds.

ACCESSION AF033353

VERSION AF033353.1 GI:2645736

KEYWORDS

SOURCE

ORGANISM

house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1187)

Howe, K., Williamson, J., Boddy, N., Sheer, D., Freemont, P. and

Solomon, E.

The ubiquitin-homology gene P1C1: characterization of mouse (P1C1)

and human (UBI1) genes and pseudogenes

Genomics 47 (1), 92-100 (1998)

2 (bases 1 to 1187)

98126440

1.1187

Location/Qualifiers

source

1.1187

/organism="Mus musculus"

/strain="ICR"

```

/db_xref="taxon:10090"
/map="C2-3"
1..1187
/gene="Ubl1"
/note="P1c1"
119..424
/gene="Ubl1"
/codon_start=1
/product="ubiquitin-homology domain protein"
/protein_id="AAC3959.1"
/db_xref="GI:2645737"
/translation="MSDDEAKPSTEDLGDKKEGEYIKLVIGDSSSEIHKVKMTTHL
KKIKESYCOGVPNMSLRFLEQRIADNHTPELGMEDYIEVYOEOTGSHSTV"
BASE COUNT      350 a      216 c      267 g      354 t
ORIGIN

alignment_scores:
  Quality: 522.00      Length: 101
  Ratio: 5.168      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.010

alignment_block:
US-09-484-964-2 x AF033353 ..

Align seg 1/1 to: AF033353 from: 1 to: 1187

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
119 ATGCTGTGACGAGGAGGCAAAACCTTCACTGAGACTTAAAGGATAGAA 168
|||||
17 sGlnGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
169 GGAAGGAGAAATACATTAACCTCAAGTATTGGACAGATGACAGTAGA 218
|||||
34 LeuHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
219 TACATTTCAAAAGTGAATAATGACACACATCTCAAGAAACCTCAAGAAATCA 268
|||||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67
|||||
269 TACTGTCAAAAGACAGGGAGTCCCATTAATTCACCTGAGGTTTCTTTGA 318
|||||
67 uGlyGlnArgGlnIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
319 AGGTCACAGAAATTCCTGATATCATCTCCGAAAGAACTGGGAATGAGG 368
|||||
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThr 100
|||||
369 AAGAAGATGTGATTGAAGTTATCAGAACAAACGGGGGCTCAGCTGAGC 418
|||||
101 Val 101
|||
419 GTT 421

seq_name: gb_pr10:HSU61397

seq_documentation_block:
LOCUS      HSU61397      1223 bp      mRNA      PRI      14-JAN-1997
DEFINITION Human ubiquitin-homology domain protein P1c1 mRNA, complete cds.
ACCESSION  U61397
VERSION     U61397.1  GI:1518693
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1223)
AUTHORS     Boddy,M.N., Howe,K., Etkin,L.D., Solomon,E. and Freemont,P.S.
TITLE       P1c1, a novel ubiquitin-like protein which interacts with the PM1
component of a multiprotein complex that is disrupted in acute
promyelocytic leukaemia

```

```

JOURNAL      Oncogene 13 (5), 971-982 (1996)
MEDLINE      96400311
REFERENCE    2 (bases 1 to 1223)
AUTHORS      Howe,K., Boddy,M.N., Etkin,D.E., Solomon,E. and Freemont,P.S.
TITLE        Direct Submission
JOURNAL      Submitted (20-JUN-1996) Somatic Cell Genetics, Imperial Cancer
Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK
FEATURES
source
1..1223
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="Placenta"
146..451
/codon_start=1
/product="ubiquitin-homology domain protein P1c1"
/protein_id="BAB40388.1"
/db_xref="GI:1518694"
/db_xref="GI:1518694"
/translation="MSDDEAKPSTEDLGDKKEGEYIKLVIGDSSSEIHKVKMTTHL
KKIKESYCOGVPNMSLRFLEQRIADNHTPELGMEDYIEVYOEOTGSHSTV"
BASE COUNT      371 a      209 c      268 g      375 t
ORIGIN

alignment_scores:
  Quality: 522.00      Length: 101
  Ratio: 5.168      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.010

alignment_block:
US-09-484-964-2 x HSU61397 ..

Align seg 1/1 to: HSU61397 from: 1 to: 1223

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
146 ATGCTGTGACGAGGAGGCAAAACCTTCACTGAGACTTGGGAGTAAAGAA 195
|||||
17 sGlnGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
196 GGAAGGTGATATTAATTAACCTCAAGTCTTGACAGATAGACAGTAGA 245
|||||
34 LeuHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
246 TACACTCAAAAGTGAATAATGACACACATCTCAAGAACTCAAGAAATCA 295
|||||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67
|||||
296 TACTGTCAAAAGACAGGGAGTCCCATTAATTCACCTGAGGTTTCTTTGA 345
|||||
67 uGlyGlnArgGlnIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
346 GGCTCAGAGAAATTCCTGATATCATCTCCAAAGAACTGGGAATGAGG 395
|||||
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThr 100
|||||
396 AAGAAGATGTGATTGAAGTTATCAGAACAAACGGGGGCTCAGCTCAACA 445
|||||
101 Val 101
|||
446 GTT 448

seq_name: gb_pr7:BC006462

seq_documentation_block:
LOCUS      BC006462      1227 bp      mRNA      PRI      13-APR-2001
DEFINITION Homo sapiens, ubiquitin-like 1 (sentrin), clone MGC:2095, mRNA,
complete cds.
ACCESSION  BC006462
VERSION     BC006462.1  GI:13623670
KEYWORDS    MGC.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE 1 (bases 1 to 1227)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Nataja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: ISAL Plate: 8 Row: f Column: 12.

FEATURES Location/Qualifiers

source

1..1227
/organism="Homo sapiens"
/db_xref="LocusID:7341"
/db_xref="taxon:9606"
/clone="MGC:2095 IMAGE:3546125"
/tissue.type="Placenta, choriocarcinoma"
/clone.lib="NIH_MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
132..437
/codon_start=1
/product="ubiquitin-like 1 (sentrin)"
/protein_id="AA06462.1"
/db_xref="GI:13623671"
/translation="MSDQKAPSTEDLDKKEGEYIKLVIGDSSSEIHFKVMTTHL
KKLSESYCORQGVPMNSLRFLEFGRIADNHTPRELGEEDVIEVYOEOTGHSYV"

CDS

BASE COUNT 386 a 208 c 259 g 374 t
ORIGIN

alignment_scores:

Quality: 522.00 Length: 101
Ratio: 5.168 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:

US-09-484-964-2 x BC006462 ..

Align seg 1/1 to: BC006462 from: 1 to: 1227

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysI 17
|||||
132 ATGCTGACCAAGGAGGCAAACTTCACAGAGACTTGGGATAGAA 181
17 sGlnGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
182 GGAAGGTGAATATTATTAAGTCAAGCATTTGGACAGGATGACAGTGA 231
34 LeuIlePheLysValLysMetThrThrHisLeuLysLysLeuLysGlu 50
|||||
232 TTCACCTCAAAAGTGAATGACACATCTCAAGAACTCAAGATCA 281

51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
282 TACGTCAAGACAGAGGTGTTCCATGAAATTCACACTGCTTCTTTGA 331
67 uGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
332 GGGTCAGAAATTCCTGATATCATCTACCAAAAGAACTGGGAATGAGG 381
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThr 100
|||||
382 AAGAAGATGATGATTGAAGTTATTCAGAAACAAACGGGGGTCATTCAACA 431
101 Val 101
|||
432 GTT 434

seq_name: gb.pat1:AR087051

seq_documentation_block:

LOCUS AR087051 1514 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 1 from patent US 5985664.
ACCESSION AR087051
VERSION AR087051.1 GI:10013817
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1514)
AUTHORS Baker, B.F. and Cowse, L.M.
TITLE Antisense modulation of Sentrin expression
JOURNAL Patent: US 5985664-A 1 16-NOV-1999;
FEATURES Location/Qualifiers

source

1..1514
/organism="unknown"
BASE COUNT 462 a 256 c 314 g 482 t
ORIGIN

alignment_scores:

Quality: 522.00 Length: 101
Ratio: 5.168 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:

US-09-484-964-2 x AR087051 ..

Align seg 1/1 to: AR087051 from: 1 to: 1514

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysI 17
|||||
136 ATGCTGACCAAGGAGGCAAACTTCACAGAGACTTGGGATAGAA 185
17 sGlnGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
186 GGAAGGTGAATATTATTAAGTCAAGTCAATTCAGACAGATGACATGAGA 235
34 LeuIlePheLysValLysMetThrThrHisLeuLysLysLeuLysGlu 50
|||||
236 TTCACCTCAAAAGTGAATGACACATCTCAAGAACTCAAGATCA 285
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
286 TACGTCAAGACAGAGGTGTTCCATGAAATTCACAGAGTTCTTTGA 335
67 uGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
336 GGGTCAGAAATTCCTGATATCATCTCAAAAGAACTGGGAATGAGG 385
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThr 100
|||||
386 AAGAAGATGATGATTGAAGTTATTCAGAAACAAACGGGGGTCATTCAACA 435
101 Val 101


```

||||
436 GTT 438

seq_name: gb_pt10:HSU83117
seq_documentation_block:
LOCUS      HSU83117      1514 bp      mRNA      PRI      09-JAN-1997
DEFINITION Human sentrin mRNA, complete cds.
ACCESSION  U83117
VERSION    U83117.1  GI:1769601
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1514)
AUTHORS   Okura,T., Gong,L., Kamitani,T., Wada,T., Okura,I., Wei,C.-F.,
            Chang,H.-M. and Yeh,E.T.H.
            Protection Against Fas/Apo-1- and Tumor Necrosis Factor-Mediated
            Cell Death by a Novel Protein, Sentrin
JOURNAL   J. Immunol. 157(10), 4277-4281 (1996)
TITLE      2 (bases 1 to 1514)
AUTHORS   Okura,T., Gong,L., Kamitani,T., Wada,T., Okura,I., Wei,C.-F.,
            Chang,H.-M. and Yeh,E.T.H.
            Direct Submission
JOURNAL   Submitted (23-DEC-1996) Division of Molecular Medicine, Department
            of Internal Medicine, and Cardiovascular Research Center, Institute
            of Molecular Medicine for the Prevention of Human Diseases, The
            University of Texas-Houston Health Science Center, 2121 W. Holcombe
            Blvd., Suite 900, Houston, TX 77030, USA
FEATURES
source     1. 1514
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /rnote="type="placenta"
            136..441
            /note="ubiquitin-like protein"
            /codon_start=1
            /product="sentrin"
            /protein_id="AA83999.1"
            /db_xref="GI:1769602"
            /translation="MSDOEAKPTEDIGCKKEGEYIKLVIGDSSSEIHPVKMTTL
            KKLKESYCORGVPMNSLRFLFGORIADNHTPELQMEEDVLEVOEOTGSHSTV"
BASE COUNT 462 a 256 c 314 g 482 t
ORIGIN
alignment_scores:
Quality: 522.00 Length: 101
Ratio: 5.168 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
US-09-484-964-2 x HSU83117 ..
Align seg 1/1 to: HSU83117 from: 1 to: 1514

1 MetsetASpGLnglGuaAlaLysProSerThrGluAspLeuGlyAspLysI 17
|||||
136 ATGTCTGACGACGAGGCAAAACCTTCACCTGAGACTGGGGATTAAGAA 185
17 sgInglGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGluI 34
|||||
186 GGAAGTGAAATATTAACCTCAAGTCATGTGACAGCATAGCAGTGA 235
34 IeHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
236 TTCACCTCAAAAGTGAAGAAATGACACACATCTCAAGAACTCAAGAAATCA 285
51 TTYCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
286 TACTGTCAAGACAGGAGGTTCCTCAATGAATCTACAGGTTTCTCTTGA 335

```

```

67 uGIglnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
|||||
336 GGCTCAGACGATTCGTGATNATCATCTCCAAAGACATGGGAATGAGG 385
386 AAGACAGATGTGATTGAAGTTTATTCAGAGAACAAACGGGGGTCTATCAACA 435
101 Val 101
||||
436 GTT 438

seq_name: gb_htg17:AC084066
seq_documentation_block:
LOCUS      AC084066      235411 bp      DNA      HTG      12-OCT-2000
DEFINITION Mus musculus clone RP23-321D1, *** SEQUENCING IN PROGRESS ***, 29
            unordered pieces.
ACCESSION  AC084066
VERSION    AC084066.1  GI:10799415
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 235411)
AUTHORS   DOE Joint Genome Institute.
TITLE      Sequencing of Mouse
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 235411)
AUTHORS   DOE Joint Genome Institute.
TITLE      DOE Joint Genome Institute.
JOURNAL   Direct Submission
            Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT    -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 2351294
            Center clone name: RPCI-23_321D1
            -----
            Summary Statistics
            Consensus quality: 214207 bases at least Q40
            Consensus quality: 223053 bases at least Q30
            Consensus quality: 225208 bases at least Q20
            Estimated insert size: 200000; pulse field gel estimation
            Estimated insert size: 232611; sum-of-contigs estimation
            Quality coverage: 11.56 in Q20 bases; pulse field gel estimation
            Quality coverage: 9.94 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            1 1210: contig of 1210 bp in length
            * 1211
            * 1310: gap of unknown length
            * 1311
            * 1311: contig of 1930 bp in length
            * 1341
            * 3241: gap of unknown length
            * 3341
            * 4504: contig of 1164 bp in length
            * 4505
            * 4604: gap of unknown length
            * 4605
            * 6172: contig of 1568 bp in length
            * 6173
            * 6272: gap of unknown length
            * 6273
            * 7354: contig of 1082 bp in length
            * 7355
            * 7454: gap of unknown length
            * 7455
            * 8625: contig of 1171 bp in length
            * 8626
            * 8725: gap of unknown length
            * 8726
            * 10114: contig of 1389 bp in length
            * 10214: gap of unknown length

```

```

* 10215 12091: contig of 1877 bp in length
* 12092 12191: gap of unknown length
* 12192 14113: contig of 1922 bp in length
* 14114 14213: gap of unknown length
* 14214 15286: contig of 1073 bp in length
* 15287 17839: gap of unknown length
* 17840 17939: contig of 2453 bp in length
* 17940 20267: gap of unknown length
* 20267 20366: contig of 2327 bp in length
* 20367 22001: gap of unknown length
* 22002 22101: contig of 1635 bp in length
* 24308 24407: gap of 2206 bp in length
* 24408 26405: gap of 1998 bp in length
* 26406 26505: gap of unknown length
* 26506 26991: contig of 3186 bp in length
* 26992 29791: gap of unknown length
* 29792 34705: contig of 4914 bp in length
* 34706 34805: gap of unknown length
* 34806 39749: contig of 4944 bp in length
* 39750 39849: gap of unknown length
* 39850 45296: contig of 3447 bp in length
* 45297 45396: gap of unknown length
* 45397 51476: contig of 6080 bp in length
* 51477 51576: gap of unknown length
* 51577 59008: contig of 7432 bp in length
* 59009 59109: gap of unknown length
* 59109 66219: contig of 7110 bp in length
* 66219 66318: gap of unknown length
* 66319 76779: gap of 10460 bp in length
* 76779 76878: gap of unknown length
* 76879 98098: contig of 21220 bp in length
* 98099 98198: gap of unknown length
* 98199 113987: contig of 15789 bp in length
* 113988 132115: gap of unknown length
* 132116 132215: contig of 18028 bp in length
* 132216 160184: gap of unknown length
* 160185 160284: gap of 27969 bp in length
* 160285 186698: contig of 26414 bp in length
* 186699 186798: gap of unknown length
* 186799 235411: contig of 48613 bp in length.
FEATURES
    source             Location/Qualifiers
    source             1..235411
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /clone_lib="RPCI mouse BAC library 23"
BASE COUNT      62132 a 54445 c 54225 g 61806 t 2803 others
ORIGIN

```

```

alignment_scores:
    Quality: 518.00      Length: 101
    Ratio: 5.129        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 98.020
alignment_block:
US-09-484-964-2 x AC084066/rev ..
Align seg 1/1 to reverse of: AC084066 from: 1 to: 235411
1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
113106 ATGTCGACCGAGGAGGCAAACTTCACAGGACTTAGCCGATAGAA 113057
|||||
17 sGlnGlyGluTyrTlleLysLeuLysValIleGlyGlnAspSerSerLui 34
|||||
113056 GGAAGGAGAAATACATTAACCTCAAGATTATGGACAGCAGCGCTGAGA 113007
|||||
34 leHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||

```

```

113006 TACATTCAAAAGTGAAATGCAACACATCTCAGAAACTCAAGATCA 112957
51 TyrCysGlnArgGlnGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67
|||||
112956 TACTGTCAAAAGCAGGAGAGTCCCAATGAATTCACAGGTTCTCTTTGA 112907
|||||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
|||||
112906 AGGTGACAGATTTGCTGATATCATCTCCGAAAGAACTGGAAATGAGA 112857
|||||
84 lUGlnAspValIleGlyValTyrGlnGlnGlnThrGlyGlyHisSerThr 100
|||||
112856 AAGAAAGATGTGATTAAGTTATACAGACAAACGCGGCTCAGTCAGC 112807
|||||
101 Val 101
112806 GTT 112804
seq_name: gb|om:AF242526
seq_documentation_block:
LOCUS      AF242526      510 bp      mRNA      NAM      03-AUG-2000
DEFINITION Cervus nippon sentrin mRNA, complete cds.
ACCESSION  AF242526
VERSION    AF242526.1 GI:9664276
KEYWORDS
SOURCE
ORGANISM
    Cervus nippon
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae;
    Cervinae; Cervus.
REFERENCE
    Sun, L.-G., Yu, Y.-L. and Jiang, Y.
    1 (bases 1 to 510)
AUTHORS
    Direct Submission
TITLE
    Submitted (07-MAR-2000) Dept. of Immunology, Norman Bethune
    University of Medical Sciences, Xin Min Street, Changchun, Jilin
    130021, People's Republic of China
FEATURES
    source             Location/Qualifiers
    source             1..510
                        /organism="Cervus nippon"
                        /db_xref="taxon:9663"
                        /note="author: Cervus nippon Temminck"
    CDS
                        129..434
                        /codon_start=1
                        /product="sentrin"
                        /protein_id="AAF97049.1"
                        /db_xref="GI:9664277"
                        /translation="MSDQKAPSTEDLDGKKEGYIKIKYIGDSSSEIHFEVKKFTTHL
                        KLRKSYCOROCVPMNSLRFLFEGQRIDNHTPKELGMEEDVLEVHEQRGHSTV"
BASE COUNT      154 a 98 c 121 g 137 t
ORIGIN

```

```

alignment_scores:
    Quality: 517.00      Length: 101
    Ratio: 5.119        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 98.020
alignment_block:
US-09-484-964-2 x AF242526 ..
Align seg 1/1 to: AF242526 from: 1 to: 510
1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
129 ATGTCGACCGAGGAGGCAAACTTCACAGGAGCTTAGCGGATAGAA 178
|||||
17 sGlnGlyGluTyrTlleLysLeuLysValIleGlyGlnAspSerSerLui 34
|||||
179 GGAAGGAGAAATATATTAACCTCAAGCTATTGGACAGGATAGCAGTGA 228
|||||
34 leHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||

```

```

229 TTCACCTCAAGATGAAATGACGACATCTCAAGAACTCAAGATCA 278
51 TYTCYSGlnArgGlnGlnValPrometAsnSerLeuArgPheLeuPheG1 67
|||||
279 TACTGTCAAGACAGGAGGTCTCTATGAAATCTCAAGGTTCTCTTTGA 328
67 uGlyGlnArgGlnGlnValPrometAsnSerLeuArgPheLeuGlyMetG1ug 84
|||||
329 AGGTCAAGATGATGCTGATATCATCACTCAAGAACTGGGATGAGG 378
84 luGluAspValIleGluValTYrGlnGlnGlnThrGlyGlnHisSerThr 100
|||||
379 AAGAAATGTGATTGAAATTATCATGACAAACAGGGGATTCATCAAG 428
101 Val 101
429 GTT 431

seq_name: gb_htg22:AL513282

seq_documentation_block:
LOCUS AL513282 155913 bp DNA HTG 26-FEB-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-179G5, *** SEQUENCING IN
PROGRESS ***, 8 unordered pieces.
ACCESSION AL513282
VERSION AL513282.5 GI:13162044
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 155913)
Pavitt,R.
Direct Submission
Submitted (25-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Feb 28, 2001 this sequence version replaced gi:13161719.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: DA17965
----- Summary Statistics
Assembly program: XGAP4; Version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 153015 bases at least Q40
Consensus quality: 154565 bases at least Q20
Insert size: 155213; sum-of-contrigs
Insert size: 164818; 1.0% error; agarose-fp
Quality coverage: 6.89x in Q20 bases; sum-of-contrigs Quality
coverage: 7.27x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20838: contrig of 20838 bp in length
* 20839 20938: gap of 100 bp
* 20939 50589: contrig of 29651 bp in length
* 50590 50689: gap of 100 bp
* 50690 81538: contrig of 30849 bp in length
* 81539 81638: gap of 100 bp
* 81639 89682: contrig of 8044 bp in length
* 89683 89782: gap of 100 bp

```

```

FEATURES
Source
1..155913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-179G5"
/clone_1fb="RP11-11.1"
1..20838
/note="assembly_fragment:02743
fragment_chain:1
clone_end:17
vector_side:left"
20939..50589
/note="assembly_fragment:01236
fragment_chain:1"
50690..81538
/note="assembly_fragment:00478
fragment_chain:2"
81639..89682
/note="assembly_fragment:03283
fragment_chain:2"
89783..108105
/note="assembly_fragment:00942
fragment_chain:3"
108206..149395
/note="assembly_fragment:02750
fragment_chain:3"
149496..153075
/note="assembly_fragment:02486
fragment_chain:4"
153176..155913
/note="assembly_fragment:03083
fragment_chain:4"
BASE COUNT 43830 a 34529 c 35326 g 41527 t 701 others
ORIGIN
alignment_scores:
Quality: 512.00 Length: 101
Ratio: 5.120 Gaps: 0
Percent Similarity: 99.010 Percent Identity: 97.030
alignment_block:
US-09-484-964-2 x AL513282/rev ..
Align seg 1/1 to reverse of: AL513282 from: 1 to: 155913
1 MetSerAspGlnGlnAlaLysProSerThrGluAspLeuGlyAspLysly 17
|||||
136335 ATGCTGACCAAGGAGGCAAACTTCACTGAGACTTGGGGATTAAGA 136306
17 sGlnGlyLysTYrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
136305 GGAAGTGAATATATTAACCTCAATGACAGATGACAGTACGAGAGA 136256
34 LeuIshPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
136255 TTCACCTCAAGATGAAATGACACACATCTCAAGAACTCAAGATCA 136206
51 TYTCYSGlnArgGlnGlnValPrometAsnSerLeuArgPheLeuPheG1 67
|||||
136205 TACTGTCAAGACAGGCGCTTCAATGAAATTCATGAGTTCTCTTTGA 136156
67 uGlyGlnArgGlnGlnValPrometAsnSerLeuArgPheLeuGlyMetG1ug 84
|||||
136155 GGGTCAGAGAAATGCTGATATCATCTCAAGAACTGGGATGAGG 136106

```

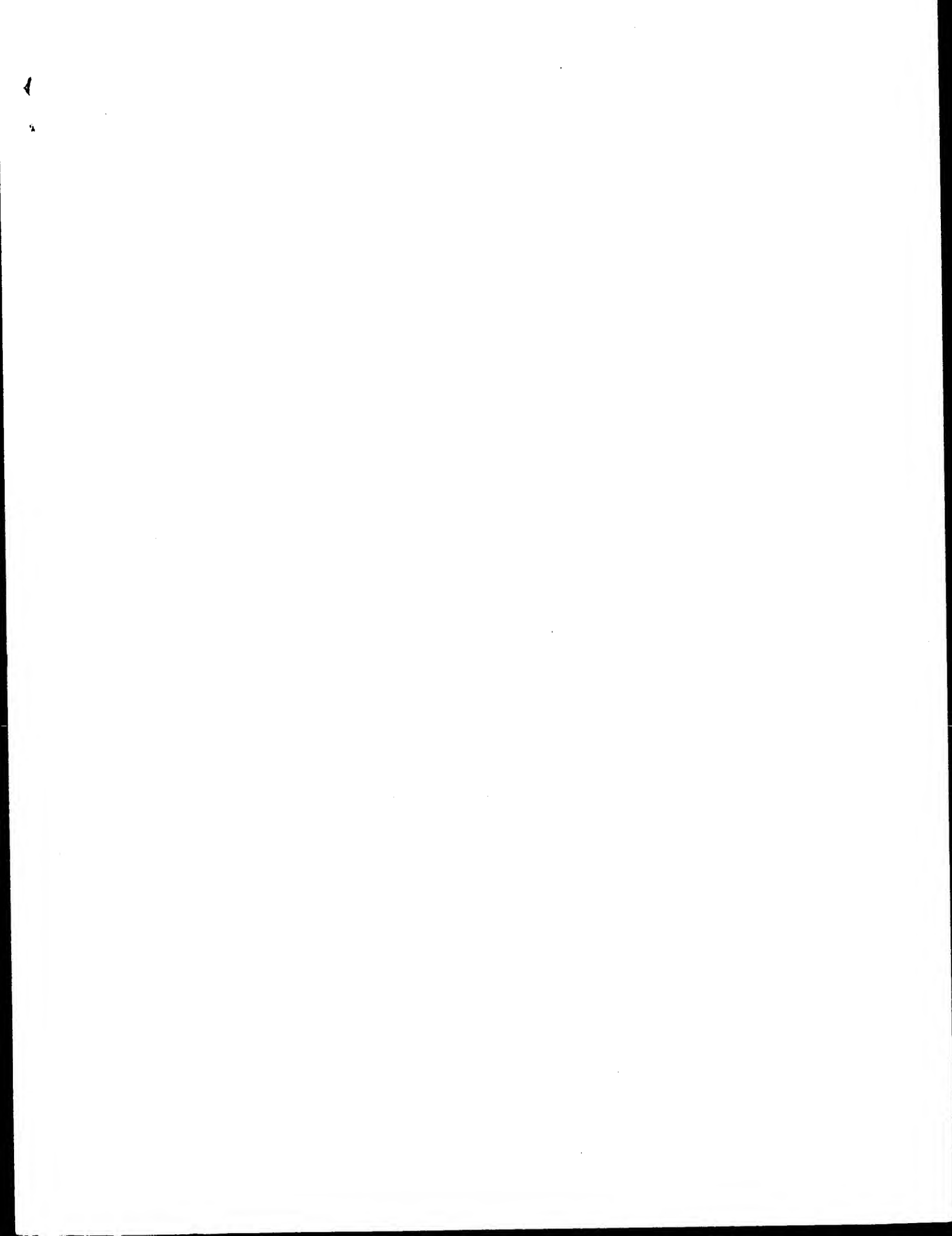
/organism="Homo sapiens"

AX021844.1 GT:10045087

Page 9

Thu Aug 16 08:48:53 2001

us-09-484-964-2.rge




```
|||||
88 ATGTCGACCGAGGCAAAACCTTCACTGAGACTTGGGGATTAAGAA 137
17 sGlnGlyLeuTyrIleLeuLeuValIleGlyLysSerSerGlu 34
138 GCAGAGTGATATATTAACTCAAGTCATTGACAGATAGCACTGACA 187
34 LeHspheLysValLysMetThrThrHisLeuLysLeuGlyLysSer 50
188 TTCACCTCAAACTGAAATGACACATCTCCAGAACTCAAGAAATCA 237
51 TTYCysGlnArgGlnGlyValProMetLysSerLeuArgPheLeuPheG 67
238 TACTGTCAAAAGCAGGGGTGTCATCAATGAACTCAGGTTCTCTTGA 287
67 uGlyLysArgIleAlaAspAsnHisThrProLysGluLeuGlyMetLug 84
288 GGGTCAGAGAAATGCTGTATATCACTACTCCAAAAGAACTGGGATGAGG 337
84 LuGlyAspValIleGlyValLysGlnGluGlnThrGlyLysSerThr 100
338 AAGAAGATGATGATTGAAGTTTATCAGCAACAAAGGGGGTCAATCAACA 387
101 Val 101
388 CTT 390

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NAL198.DAT:AAV59722
seq_documentation_block:
ID AAV59722 standard; DNA; 1196 BP.
XX
AC AAV59722;
XX
DT 19-JAN-1999 (first entry)
XX
DE Human secreted protein gene 47 clone HOGAV75.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO839448-A2.
XX
PD 11-SEP-1998.
XX
PE 06-MAR-1998: 98WO-US04493.
XX
XX 02-OCT-1997: 97US-0061060.
PR 07-MAR-1997: 97US-0038621.
PR 07-MAR-1997: 97US-0040161.
PR 07-MAR-1997: 97US-0040162.
PR 07-MAR-1997: 97US-0040163.
PR 07-MAR-1997: 97US-0040333.
PR 07-MAR-1997: 97US-0040334.
PR 07-MAR-1997: 97US-0040336.
PR 07-MAR-1997: 97US-0040626.
PR 11-APR-1997: 97US-0043311.
PR 11-APR-1997: 97US-0043312.
PR 11-APR-1997: 97US-0043313.
PR 11-APR-1997: 97US-0043314.
PR 11-APR-1997: 97US-0043568.
PR 11-APR-1997: 97US-0043569.
PR 11-APR-1997: 97US-0043576.
PR 11-APR-1997: 97US-0043578.
PR 11-APR-1997: 97US-0043580.
```

```
PR 11-APR-1997: 97US-0043669.
PR 11-APR-1997: 97US-0043670.
PR 11-APR-1997: 97US-0043671.
PR 11-APR-1997: 97US-0043672.
PR 23-MAY-1997: 97US-0043674.
PR 23-MAY-1997: 97US-0047492.
PR 23-MAY-1997: 97US-0047500.
PR 23-MAY-1997: 97US-0047501.
PR 23-MAY-1997: 97US-0047502.
PR 23-MAY-1997: 97US-0047503.
PR 23-MAY-1997: 97US-0047581.
PR 23-MAY-1997: 97US-0047582.
PR 23-MAY-1997: 97US-0047583.
PR 23-MAY-1997: 97US-0047584.
PR 23-MAY-1997: 97US-0047585.
PR 23-MAY-1997: 97US-0047586.
PR 23-MAY-1997: 97US-0047587.
PR 23-MAY-1997: 97US-0047588.
PR 23-MAY-1997: 97US-0047589.
PR 23-MAY-1997: 97US-0047590.
PR 23-MAY-1997: 97US-0047592.
PR 23-MAY-1997: 97US-0047593.
PR 23-MAY-1997: 97US-0047594.
PR 23-MAY-1997: 97US-0047595.
PR 23-MAY-1997: 97US-0047596.
PR 23-MAY-1997: 97US-0047597.
PR 23-MAY-1997: 97US-0047598.
PR 23-MAY-1997: 97US-0047599.
PR 23-MAY-1997: 97US-0047600.
PR 23-MAY-1997: 97US-0047601.
PR 23-MAY-1997: 97US-0047612.
PR 23-MAY-1997: 97US-0047613.
PR 23-MAY-1997: 97US-0047614.
PR 23-MAY-1997: 97US-0047615.
PR 23-MAY-1997: 97US-0047617.
PR 23-MAY-1997: 97US-0047618.
PR 23-MAY-1997: 97US-0047632.
PR 23-MAY-1997: 97US-0047633.
PR 06-JUN-1997: 97US-0048974.
PR 06-JUN-1997: 97US-0048976.
PR 13-JUN-1997: 97US-0048977.
PR 08-JUL-1997: 97US-0049610.
PR 16-JUL-1997: 97US-0051926.
PR 18-AUG-1997: 97US-0052874.
PR 22-AUG-1997: 97US-0052874.
PR 22-AUG-1997: 97US-0056630.
PR 22-AUG-1997: 97US-0056631.
PR 22-AUG-1997: 97US-0056632.
PR 22-AUG-1997: 97US-0056636.
PR 22-AUG-1997: 97US-0056637.
PR 22-AUG-1997: 97US-0056662.
PR 22-AUG-1997: 97US-0056664.
PR 22-AUG-1997: 97US-0056845.
PR 22-AUG-1997: 97US-0056846.
PR 22-AUG-1997: 97US-0056862.
PR 22-AUG-1997: 97US-0056864.
PR 22-AUG-1997: 97US-0056872.
PR 22-AUG-1997: 97US-0056874.
PR 22-AUG-1997: 97US-0056875.
PR 22-AUG-1997: 97US-0056876.
PR 22-AUG-1997: 97US-0056877.
PR 22-AUG-1997: 97US-0056878.
PR 22-AUG-1997: 97US-0056879.
PR 22-AUG-1997: 97US-0056880.
PR 22-AUG-1997: 97US-0056881.
PR 22-AUG-1997: 97US-0056882.
PR 22-AUG-1997: 97US-0056884.
PR 22-AUG-1997: 97US-0056886.
PR 22-AUG-1997: 97US-0056887.
PR 22-AUG-1997: 97US-0056888.
PR 22-AUG-1997: 97US-0056889.
PR 22-AUG-1997: 97US-0056892.
PR 22-AUG-1997: 97US-0056893.
PR 22-AUG-1997: 97US-0056894.
PR 22-AUG-1997: 97US-0056903.
```

PR 22-AUG-1997; 97US-0056908.
 PR 22-AUG-1997; 97US-0056909.
 PR 22-AUG-1997; 97US-0056910.
 PR 22-AUG-1997; 97US-0056911.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057659.
 PR 05-SEP-1997; 97US-0057761.
 PR 12-SEP-1997; 97US-0058785.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
 PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
 PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 XX

DR MPI, 1998-506364/43.
 DR P-PDB; AAW74938.

XX New isolated human genes and the secreted polypeptide(s) they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 PS Claim 1, Page 449; 721pp; English.

XX This sequence represents a nucleic acid molecule designated Gene 47 from
 CC the human cDNA clone HOGAV75 (deposited as clone ATCC 97897 and ATCC
 CC 209043) which encodes a secreted human protein. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
 CC the fused protein as compared to the human protein only.
 CC The invention relates to 186 novel genes and their fragments (nucleic
 CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 186
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAV59511 for described uses).

XX Sequence 1196 BP; 360 A; 212 C; 259 G; 364 T; 1 other;

alignment_scores:

Quality: 522.00 Length: 101
 Ratio: 5.168 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:

US-09-484-964-2 x AAV59722 ..

Align seg 1/1 to: AAV59722 from: 1 to: 1196

1 MetSerAspGlnGlnAlaIysProSerThrGluAspLeuGlyAspIysIy 17
 134 ATGTCGTACACAGAGGCAAACTTCACAGAGCTGGGGATAGAA 183
 17 sgInglyGluTyrIleLeuLeuIleGlyGlnAspSerSerGluI 34
 184 GGAAGGGAATATTAATCAAGTCATTCAGACAGGATGACAGAGAGA 233
 34 LeuIlePheValIysMetThrThrIleLeuIleLeuIleLeuIle 50
 234 TTCACCTCAAGTGAATGACACACATCTCAAGAACTCAAGATCA 283
 51 TyrGlnGlnArgGlnGlnGlyValPrometAsnSerLeuArgPheLeuPheG 67
 284 TACTGTCAAGACAGAGGTTCATCAATGATTCACAGCTTCCTTGA 333
 67 uGlyGlnArgIleAlaAspAsnHisThrProIysGlnLeuGlyMetGluG 84
 334 GGGTCAGAGAAATTCGATATATCATCTCCAAAGAACTGGCAATGAGAG 383

84 LuGluAspValIleGlyValTyrGlnGlnGlnThrGlyIyHisSerThr 100
 384 AAGAGATGATGATTGAATTCACAGAAACAAACGGGGGTTCATTCACA 433
 101 Val 101
 434 GTT 436

seq_name: /SIDSL/gcdata/geneseq/geneseqn/NA1998.DAT:AAV59557

seq_documentation_block:

ID AAV59557 standard; DNA; 1220 BP.

XX AAV59557;

DT 06-JAN-1999 (first entry)

DE Human secreted protein gene 47 clone HOGAV75.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

PN W09839448-A2.

XX 11-SEP-1998.

PF 06-MAR-1998; 98WO-US04493.

XX 02-OCT-1997; 97US-0061060.
 PR 07-MAR-1997; 97US-0038621.
 PR 07-MAR-1997; 97US-0040161.
 PR 07-MAR-1997; 97US-0040162.
 PR 07-MAR-1997; 97US-0040163.
 PR 07-MAR-1997; 97US-0040333.
 PR 07-MAR-1997; 97US-0040334.
 PR 07-MAR-1997; 97US-0040336.
 PR 07-MAR-1997; 97US-0040626.
 PR 11-APR-1997; 97US-0043311.
 PR 11-APR-1997; 97US-0043312.
 PR 11-APR-1997; 97US-0043313.
 PR 11-APR-1997; 97US-0043314.
 PR 11-APR-1997; 97US-0043568.
 PR 11-APR-1997; 97US-0043569.
 PR 11-APR-1997; 97US-0043576.
 PR 11-APR-1997; 97US-0043578.
 PR 11-APR-1997; 97US-0043580.
 PR 11-APR-1997; 97US-0043669.
 PR 11-APR-1997; 97US-0043670.
 PR 11-APR-1997; 97US-0043671.
 PR 11-APR-1997; 97US-0043672.
 PR 11-APR-1997; 97US-0043674.
 PR 23-MAY-1997; 97US-0047492.
 PR 23-MAY-1997; 97US-0047500.
 PR 23-MAY-1997; 97US-0047501.
 PR 23-MAY-1997; 97US-0047502.
 PR 23-MAY-1997; 97US-0047503.
 PR 23-MAY-1997; 97US-0047581.
 PR 23-MAY-1997; 97US-0047582.
 PR 23-MAY-1997; 97US-0047583.
 PR 23-MAY-1997; 97US-0047584.
 PR 23-MAY-1997; 97US-0047585.
 PR 23-MAY-1997; 97US-0047586.
 PR 23-MAY-1997; 97US-0047587.
 PR 23-MAY-1997; 97US-0047588.

PR 23-MAY-1997; 97US-0047589.
 PR 23-MAY-1997; 97US-0047590.
 PR 23-MAY-1997; 97US-0047592.
 PR 23-MAY-1997; 97US-0047593.
 PR 23-MAY-1997; 97US-0047594.
 PR 23-MAY-1997; 97US-0047595.
 PR 23-MAY-1997; 97US-0047596.
 PR 23-MAY-1997; 97US-0047597.
 PR 23-MAY-1997; 97US-0047598.
 PR 23-MAY-1997; 97US-0047599.
 PR 23-MAY-1997; 97US-0047600.
 PR 23-MAY-1997; 97US-0047601.
 PR 23-MAY-1997; 97US-0047612.
 PR 23-MAY-1997; 97US-0047613.
 PR 23-MAY-1997; 97US-0047614.
 PR 23-MAY-1997; 97US-0047615.
 PR 23-MAY-1997; 97US-0047617.
 PR 23-MAY-1997; 97US-0047618.
 PR 23-MAY-1997; 97US-0047619.
 PR 23-MAY-1997; 97US-0047632.
 PR 23-MAY-1997; 97US-0047633.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048974.
 PR 13-JUN-1997; 97US-0049610.
 PR 08-JUL-1997; 97US-0051926.
 PR 16-JUL-1997; 97US-0052874.
 PR 18-AUG-1997; 97US-0055724.
 PR 22-AUG-1997; 97US-0056630.
 PR 22-AUG-1997; 97US-0056631.
 PR 22-AUG-1997; 97US-0056632.
 PR 22-AUG-1997; 97US-0056636.
 PR 22-AUG-1997; 97US-0056637.
 PR 22-AUG-1997; 97US-0056662.
 PR 22-AUG-1997; 97US-0056664.
 PR 22-AUG-1997; 97US-0056845.
 PR 22-AUG-1997; 97US-0056862.
 PR 22-AUG-1997; 97US-0056864.
 PR 22-AUG-1997; 97US-0056872.
 PR 22-AUG-1997; 97US-0056874.
 PR 22-AUG-1997; 97US-0056875.
 PR 22-AUG-1997; 97US-0056876.
 PR 22-AUG-1997; 97US-0056877.
 PR 22-AUG-1997; 97US-0056878.
 PR 22-AUG-1997; 97US-0056879.
 PR 22-AUG-1997; 97US-0056880.
 PR 22-AUG-1997; 97US-0056881.
 PR 22-AUG-1997; 97US-0056882.
 PR 22-AUG-1997; 97US-0056884.
 PR 22-AUG-1997; 97US-0056886.
 PR 22-AUG-1997; 97US-0056887.
 PR 22-AUG-1997; 97US-0056888.
 PR 22-AUG-1997; 97US-0056889.
 PR 22-AUG-1997; 97US-0056892.
 PR 22-AUG-1997; 97US-0056893.
 PR 22-AUG-1997; 97US-0056894.
 PR 22-AUG-1997; 97US-0056903.
 PR 22-AUG-1997; 97US-0056908.
 PR 22-AUG-1997; 97US-0056909.
 PR 22-AUG-1997; 97US-0056910.
 PR 22-AUG-1997; 97US-0056911.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057659.
 PR 05-SEP-1997; 97US-0057761.
 PR 12-SEP-1997; 97US-0058785.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
 PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 XX WPI, 1998-506364/43.
 DR P-PSDB: AAW74777.

XX New isolated human genes and the secreted polypeptide(s) they encode
 PR - useful for diagnosis and treatment of e.g. cancers, neurological
 PR disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1: Page 279-280; 721pp; English.
 XX
 CC This sequence represents a nucleic acid molecule designated Gene 47 from
 CC the human CDNA clone HOGAV75 (deposited as clone ATCC 97899 and ATCC
 CC 209045) which encodes a secreted human protein. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
 CC the fused protein as compared to the human protein only.
 CC The invention relates to 186 novel genes and their fragments (nucleic
 CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological conditions
 CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Specific uses are described for each of the 186
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAV59511 for described uses).
 XX
 SQ Sequence 1220 BP; 382 A; 207 C; 258 G; 372 T; 1 other;

alignment_scores: length: 101
 Quality: 522.00 Gaps: 0
 Ratio: 5.168
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
 US-09-484-964-2 x AAV59557 ..

Align seq 1/1 to: AAV59557 from: 1 to: 1220

1 MetSerAspGlnGluAlaLysProSerThrGluAspGluLysPlySly 17
 128 ATGCTGACCAAGGCAAAACCTTCACTGAGACTGGGGGTAAGAA 177
 17 GlnGlyGlyTyrTleLysLeuLysValIleGlyGlnAspSerSertLut 34
 178 GGAAGTGAATATATTAACTCAAGTCATTGGACAGATAGCAGTGA 227
 34 IeHisPheLysValLysMetThrThrHisLeuLysLysLysGluSer 50
 228 TTCACCTCAAGTGAATATGACAAACATCTCAAGAAACCTCAAGATCA 277
 51 TyrCysGlnArgGlnGlyValPrometasSerLeuArgPheLeuPheG 67
 278 TACGTCAAGACAGAGGCTGTCCAAATCAATCACTCAGTTCTCTTGA 327
 67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
 328 GGGTCAGAGAAATTCGATATATCTACTCCAAAGAACTGGAAATGAGG 377
 84 IugLysPheValIleGluValTyrGlnGluGlnThrGlyLysHisSerThr 100
 378 AAGGAGATGATTTGAAGTTATCAGGACAAACGGGGGCTCATTCACA 427
 101 Val 101
 428 GTT 430

seq_name: /STDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ35859

seq_documentation_block:

ID AAZ35859 standard; cDNA; 1514 BP.

AC AAZ35859;

DT 03-FEB-2000 (first entry)

DE Human sentrin nucleotide sequence.
 XX
 KW Human; sentrin; antisense oligonucleotide; phosphorothioate;
 XX inhibition; modulation; expression; diagnosis; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 136..441
 FT /tag= a
 FT /product= "sentrin"
 XX
 XX US5985664-A.
 XX
 XX 16-NOV-1999.
 XX
 XX 17-DEC-1998; 98US-0213768.
 XX
 XX 17-DEC-1998; 98US-0213768.
 XX
 XX (ISIS-) ISIS PHARM INC.
 XX Baker BF, Cowser LM;
 XX
 XX WPI: 2000-022284/02.
 XX P-PSDB: AAY49967.
 DR
 DR Antisense compound which modulates human sentrin expression, useful for
 PT treating diseases associated with sentrin expression -
 XX
 XX Example 13; Column 39-42; 29pp; English.
 PS
 XX The present invention describes an antisense compound (I) 8-30
 CC nucleotides long targeted to a nucleic acid molecule encoding human
 CC sentrin. The antisense compound comprises a phosphorothioate antisense
 CC oligonucleotide which inhibits expression of human sentrin. (I) is
 CC useful for inhibiting expression of sentrin in human cells or tissues
 CC in vitro, for treating humans or other animals suspected of having or
 CC being prone to a disease associated with sentrin expression. (I) can
 CC also be used for research or diagnostic purposes. The present
 CC sequence encodes human sentrin.
 CC
 SO Sequence 1514 BP; 462 A; 256 C; 314 G; 482 T; 0 other;

alignment_scores:
 Quality: 522.00 Length: 101
 Ratio: 5.168 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
 US-09-484-964-2 x AA235859 ..

Align seg 1/1 to: AA235859 from: 1 to: 1514

```

1 MeterspGingluAlaLysProSerThGluaspLeuclgLySly 17
|||||
136 ATGCTGCGCAGGAGCAAACTTCAACTGAGGACTGGGGGATAAGAA 185
17 sGlgLgLuTfTlLysLeuLysValLgGlgLaspSerSerLui 34
|:::|
186 GGAAGGGAATATATTAACCTCAAGTCATGGACAGATAGCGTGA 235
34 LHisPhelysValLysMetThrThRHisLeuLysLysLeuLysGluSer 50
|||||
236 TTCACCTCAAGTGAATGCAACACATCTCCAGAAACTCAAGATCA 285
51 TTrcysGlnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheG1 67
|||||
286 TACTGTCAAAAGACAGGCGTGTCCAAATGAAATCACTCAGGTTTCTTGA 335
67 uGlgLArgTglLeaLaspAsnH1stHProLysGluLeuGlyMetLug 84
|||||

```

336 GGCTCAGAGATTGCTGATATCATCTCCAAAGAACTGGGATGAGG 385
 84 luGluaspVallllegluValTyrGlnGluGlnThGlyGlyHisSerThr 100
 |||||
 386 AAGAAAGATGTGATTGAAGTTCATCAGGAACAAACGGGGCTCATTCACA 435
 101 Val 101
 |||
 436 GTT 438

seq_name: /SID51/gcdata/geneseq/geneseqn/NA2000.DAT:AAA16215

seq_documentation_block:
 ID AAA16215 standard; DNA; 616 BP.
 XX
 AC AAA16215;
 XX
 DT 14-JUN-2000 (first entry)
 XX
 DE Human colon cancer differentially expressed nucleotide sequence #220.
 XX
 KW Colon cancer; detect; differential expression; human; treatment;
 KW detect mutation; non-invasive diagnostic method; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200012702-A2.
 XX
 PD 09-MAR-2000.
 XX
 PE 30-AUG-1999; 99WO-US19424.
 XX
 PR 31-AUG-1998; 98US-0098639.
 PR 27-JAN-1999; 99US-0117393.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Endge WO, Steimann KE, Asle JH, Burgess CC, Carroll E;
 PI Catino TJ, Divedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
 PI Schlegel R;
 XX
 DR WPI: 2000-256641/22.
 XX
 XX Novel nucleic acids and proteins for identifying therapeutic agents
 PT useful for treating and diagnosing cancer, especially colon cancer -
 PS
 PS Claim 16; Page 211-212; 345pp; English.
 CC This sequence represents a human nucleotide sequence which is
 CC differentially expressed in colon cancer cells compared to the expression
 CC levels in normal cells. The nucleotide sequence can be used as a source
 CC of primers and probes. The nucleotide sequence is useful for determining
 CC the phenotype of a cell by detecting the differential expression of the
 CC sequence relative to a normal cell. The probes derived from the sequence
 CC can also be used to determine the phenotype of cells in a sample. Probes
 CC and antibodies which hybridise to the nucleotide sequence can also be
 CC used to determine the phenotype of a cell. The primers are useful for
 CC detecting a mutation in a test nucleotide sequence and also for detecting
 CC cancer, preferably colon cancer. Antibodies against the protein encoded
 CC by the nucleotide sequence can also be used in a method to detect colon
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
 CC colon cancer at an early stage.
 CC
 SO Sequence 616 BP; 181 A; 109 C; 149 G; 170 T; 7 other;

alignment_scores:
 Quality: 503.00 Length: 102
 Ratio: 5.030 Gaps: 1
 Percent Similarity: 98.039 Percent Identity: 97.059

alignment_block:
 US-09-484-964-2 x AAA16215 ..

Align seg 1/1 to: AAA16215 from: 1 to: 616

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
127 ATGCTGACAGAGGAGGCAAAACCTTCAACTGAGAGCTTGGGGATTAAGAA 176
  |||
17 sGlnGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
  |||
177 GGAAGGCAATATTTAACTCAAAAGTCATTGGACAGATTCAGCTGAGA 226
  |||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLysGluSer 50
  |||
227 TTCACCTCAAAAGTCAAAATGACACACATCTCAAGAAACTCAAAAGATCA 276
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
  |||
277 TACGTCTCAAGACAGGGGTTCACATGATTCACAGCTTCTCTTGA 326
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
  |||
327 GGGTCACAGAAATTCGTATATCATCTACTCCAAAGAACTGGGAATGAGG 376
  |||
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerTh 100
  |||
377 AAGAAGATGTGATTGAAGATTATCANGAACAAACCGGGGTCTATTCAC 426
  |||
100 rVal 101
  |||
427 AGTT 430

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:AA163339

seq_documentation_block:

ID AAT63339 standard; CDNA; 372 BP.

AA163339;

17-AUG-1997 (first entry)

Human host cell protein NPI-6 partial cDNA clone.

Host cell protein; NPI-6; nucleoprotein interactor 6;

Influenza virus; replication; antiviral; virucide; ss.

Homo sapiens.

W09712967-A1.

10-APR-1997.

06-OCT-1995; 95WO-US13044.

06-OCT-1995; 95WO-US13044.

(MOON) MOUNT SINAI MEDICAL CENT.

Oneill R, Palese P;

WPI; 1997-226211/20.

New isolated DNA which encodes viral interacting proteins - used in assays to isolate products for inhibiting viral protein binding which is required for infection, replication, assembly or release

Disclowse; Fig 11; 98pp; English.

CDNA clones (AAT63335-39) comprise partial sequences for human

nucleoprotein interactor proteins NPI-2 to NPI-6, respectively.

These are host cell proteins which interact with influenza virus

nucleoprotein (NP) and which may be accessory proteins required for

influenza virus replication. NPI-6 was identified as a novel

protein sequence. NPI sequences were isolated by interactive trap

CC selection using LexA-NP as bait and yeast transformed with an HeLa
CC CDNA library. NPI-1 (see also AAT63334) and NPI-2 to NPI-6 cDNAs were
CC cloned and analysed. The clones can be used to produce human NPI
CC proteins for use in identifying cpds. that inhibit viral
CC replication.

SQ Sequence 372 BP; 121 A; 72 C; 97 G; 82 T; 0 other;

alignment_scores:

Quality: 492.00 Length: 97

Ratio: 5.125 Gaps: 0

Percent Similarity: 98.969 Percent Identity: 97.938

alignment_block:

US-09-484-964-2 x AAT63339 ..

Align seg 1/1 to: AAT63339 from: 1 to: 372

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
80 ATGCTGACAGAGGAGGCAAAACCTTCAACTGAGAGCTTGGGGATTAAGAA 129
  |||
17 sGlnGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGlu 34
  |||
130 GGAAGGCAATATTTAACTCAAAAGTCATTGGACAGATTCAGCTGAGA 179
  |||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLysGluSer 50
  |||
180 TTCACCTCAAAAGTCAAAATGACACACATCTCAAGAAACTCAAAAGATCA 229
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
  |||
230 TACGTCTCAAGACAGGGGTTCACATGATTCACAGCTTCTCTTGA 279
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
  |||
280 GGGTCACAGAAATTCGTATATCATCTACTCCAAAGAACTGGGAATGAGG 329
  |||
84 LuGluAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerTh 97
  |||
330 AAGAAGATGTGATTGAAGATTATCANGAACAAACCGGGGTCTATTCAC

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF59399

seq_documentation_block:

ID AAF59399 standard; DNA; 372 BP.

AAF59399;

02-MAY-2001 (first entry)

Human host cell protein NPI-6 partial nucleotide sequence SEQ ID NO:11.

Identification; antiviral; viral protein; viral replication; NP;

viral infection; nucleoprotein; ds.

Homo sapiens.

W0200111335-A2.

15-FEB-2001.

11-AUG-2000; 2000MO-US22257.

11-AUG-1999; 99US-0148263.

(MOON) MOUNT SINAI SCHOOL MEDICINE.

O'Neill R, Harty R, Palese PW;

WPI; 2001-168816/17.

PT Identifying a substance that inhibits the interaction between a viral
 PT protein and a host cell protein, useful for the discovery of new
 PT antiviral compounds -

PS Disclosure; Fig 11; 147pp; English.

CC The present invention describes a method (M1) for identifying a
 CC substance that inhibits the interaction of a viral protein (VP) with a
 CC host cell protein (HP). The method comprises: (a) contacting HP with VP
 CC in the presence of a test substance; and (b) detecting complex formation,
 CC where the ability of the test substance to inhibit HP/VP interaction is
 CC indicated by a decrease in complex formation. The antiviral compounds
 CC that inhibit the interaction between a host protein (NS1-BP or NP1-1)
 CC and a viral protein (NS1) are useful for treating or inhibiting viral
 CC infection, preferably influenza and rhinovirus infection, in humans.
 CC Antiviral compounds include peptides and antibodies. In particular
 CC compositions comprising a polypeptide containing an amino acid sequence
 CC corresponding to the NP-NS1 domain of the influenza virus NP protein,
 CC which inhibits the specific interaction of the influenza virus NP protein,
 CC influenza virus NP protein are useful for treating or inhibiting
 CC human host cell infection in humans. The present sequence represents a
 CC human host cell protein NP1-6 partial nucleotide sequence, which is used
 CC in the exemplification of the present invention.

SQ Sequence 372 BP; 121 A; 72 C; 97 G; 82 T; 0 other;

Alignment_scores:

Quality: 492.00 Length: 97
 Ratio: 5.125 Gaps: 0
 Percent Similarity: 98.969 Percent Identity: 97.938

alignment_block:

US-09-484-964-2 x AAF59399 ..

Align seg 1/1 to: AAF59399 from: 1 to: 372

```

1 MetSerAspGlnGlnAlaLysProSerThrGluAspLeuGlyAspLysLy 17
  |||||
80 ATGCTGACACAGAGCAAAACCTTCACACTGAGGAGCTGGGGGATAGAA 129
  |||||
17 sGlnGlyGlyTrpIleLysLeuLysValIleGlyGlnAspSerSerglu 34
  |||||
130 GGAAGGTGAATATATTAACCTCAAGCATGACAGAGATAGCAGTGAGA 179
  |||||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuGlnSer 50
  |||||
180 TTCACCTCAAGTGAATGACAAACACATCTCAAGAACTCAAGAAATCA 229
  |||||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPhe 67
  |||||
230 TACTGTCAAAAGACAGGGGTGTCATGATGATCACTCAAGTTCTCTTGA 279
  |||||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGlnLeuGlyMetGlu 84
  |||||
280 GGGTCAGAGAAATTCGTATATCACTCCAAAAGAACTGGGAATGAGG 329
  |||||
84 LuGluAspValIleGluValTyrGlnGlnGluInThrGlyGly 97
  |||||
330 AAGAAGTTGATTCAGACTTATACGAAACAAACGGGGGCT 370
  |||||

```

seq_name: /STDS1/gcgdata/geneseq/geneseqn/NA1996.DAT: AAT03735

seq_documentation_block:

ID AAT03735 standard; cDNA: 425 BP.

AC AAT03735;

DT 26-MAR-1996 (first entry)

DE TNF-R p55IC/Fas-IC-binding protein DD11 cDNA clone.

XX Tumour necrosis factor receptor; TNF-R; p55IC; Fas-IC;

KW intracellular domain binding protein; ss.

XX Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

SQ Sequence 425 BP; 135 A; 80 C; 115 G; 95 T; 0 other;

Alignment_scores:

Quality: 470.50 Length: 102
 Ratio: 4.801 Gaps: 2
 Percent Similarity: 96.078 Percent Identity: 91.176

alignment_block:

US-09-484-964-2 x AAT03735 ..

Align seg 1/1 to: AAT03735 from: 1 to: 425

```

1 MetSerAspGlnGlnAlaLysProSerThrGluAspLeuGlyAspLysLy 17
  |||||
118 ATGCTGACACAGAGCAAAACCTTCACACTGAGGAGCTGGGGGATAGAA 167
  |||||
17 ySglnGlyGlyTrpIleLysLeuLysValIleGlyGlnAspSerSerglu 33
  |||||
168 AGAAGGTGAATATATTAACCTCAAGCATGACAGAGATAGCAGTGAGA 217
  |||||
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuGlnSer 50
  |||||
218 ATTCACCTCAAGTGAATGACAAACACATCTCAAGAACTCAAGAAATC 267
  |||||
50 rTyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPhe 67
  |||||
268 ATACTGTCAAAAGACAGGGGTGTCATGATGATCACTCAAGTTCTCTTGA 317
  |||||
67 LuGlyGlnArgIleAlaAspAsnHisThrProLysGlnLeuGlyMetGlu 83
  |||||
318 AGGTCAGAGAAATTCGTATATCACTCCAAAAGAACTGGGAATGAGG 367
  |||||
84 GluGluAspVal...IleGluValTyrGlnGlnGluInThrGlyHis 99
  |||||
368 AAGAAGTTGATTCAGACTTATACGAAACAAACGGGGGCTTCATTC 417
  |||||
99 rThr 100
  |||||

```

418 AACA 421

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA43182

seq_documentation_block:

ID AAA43182 standard; CDNA; 335 BP.

AC AAA43182;

XX 21-AUG-2000 (first entry)

XX Xenopus secreted expressed sequence tag SEQ ID NO:1922.

Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
 expressed sequence tag; EST; probe; chemotactic; proliferative;
 immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
 thrombolytic; antiinflammatory; cytosolic; antibacterial; antifungal;
 antiviral; antidiabetic; neuroprotective; nociceptive; antiparasitic;
 antitumor; osteoprotective; anticonvulsant; antidepressant; gene therapy;
 cerebroprotective; anticonvulsant; multiple sclerosis; allergic condition;
 insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 central nervous system disorder; Alzheimer's disease; stroke;
 Parkinson's disease; Huntington's disease; coagulation disorder;
 haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 tumour; infection; depression; psoriasis; ss.

XX Xenopus sp.

XX WO200021990-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99MO-US24205.

XX 15-OCT-1998; 98US-0104435.

XX (GEM) GENETICS INST INC.

XX Jacobs K, McCoy JM, LaValle ER, Collins-Racie LA, Evans C;

XX Merberg D, Treacy M;

XX WPI; 2000-317937/27.

Isolated polynucleotides, and encoded proteins, comprising secreted
 expressed sequence tags (SESTs), useful for treating various disorders
 such as autoimmune, infectious, and central nervous system disorders -
 Claim 1; Page 559; 618pp; English.

AAA41261 to AAA43419 represent specifically claimed secreted expressed
 sequence tags (SESTs), isolated from human, mouse, xenopus and rat
 tissue sources. The SESTs can have a range of activities depending on
 the tissues they were isolated from. The activities include:
 chemotactic; proliferative; immunomodulatory; haematopoietic;
 chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;
 cytosolic; antibacterial; antifungal; antiviral; antidiabetic;
 antitumor; osteoprotective; anticonvulsant; antidepressant;
 nociceptive; antiparasitic; neuroprotective;
 anticonvulsant; and antidiabetic. The SESTs can be used for gene
 therapy and in vaccines. The SESTs are useful as probes for the
 identification and isolation of full-length cDNAs and genomic DNA
 molecules which correspond to the SESTs. Proteins encoded by the SESTs
 are useful in assays for determining biological activity and raising
 antibodies. They may be useful for treatment of autoimmune disorders
 (multiple sclerosis, insulin dependent diabetes), allergic conditions
 (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 osteoporosis, osteoarthritis, central nervous system disorders
 (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 disease), tumours, bacterial, fungal or viral infections, depression and
 psoriasis. AAA43420 to AAA43425 represent linker variants which are given

CC in the exemplification of the present invention.

XX Sequence 335 BP; 117 A; 61 C; 86 G; 71 T; 0 Other;

SQ

alignment_scores:

Quality: 452.50 Length: 98

Ratio: 4.714 Gaps: 1

Percent Similarity: 97.959 Percent Identity: 90.816

alignment_block:

US-09-484-964-2 x AAA43182

Align seg 1/1 to: AAA43182 from: 1 to: 335

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLy 17
 |||||
 33 ATGTCGTGATCAGAAAGCGAACCATCTAGCGAGATCTAGGACAAAGAA 82
 17 SGLN... GlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerg 33
 |||||
 83 AGATGAGGGGATTTATCAAACTAAAGTCATTGACAGACAGACAGT 132
 |||||
 33 IuLLeHIsPhelYsValLysMetThrThrHisLeuLysLysLeuLysGlu 49
 |||||
 133 AAATTCACCTTCAAGGTAGACATGACACGACATCTCAAAAAGCTGAAG 182
 50 SerTYrCYsGlnArgGlnGlyAlaProMetLysSerLeuArgPheLeuPh 66
 |||||
 183 TCATACCTCTCAGACAGCGCGCTCCAAATGAAATCTCTCAGGTTTGT 232
 66 eGluGluGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetG 83
 |||||
 233 TGAAGGCGAAAGATCTCAGATCACCAGACGCTCTAAGAGCTCGAATGG 282
 83 IuGluGluAspValIleGluValTYrGlnGlnGlnThrGlyGly 97
 |||||
 283 ACGAAGAGGATGTTATTTGAAGTTATTCAGAAACAGACACTGTGGGT 326

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC34206

seq_documentation_block:

ID AAC34206 standard; DNA; 710 BP.

XX AAC34206;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 5835.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145228.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151438.
PR 31-AUG-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154019.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

seq_documentation_block:
ID AAC35316 standard; DNA; 643 bp.
AC AAC35316;
XX
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9746.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143942.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

```

PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152366.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores:

Quality:	222.00	Length:	94
Ratio:	3.083	Gaps:	1
Percent Similarity:	76.596	Percent Identity:	48.936

alignment_block:

US-09-484-964-2 x AAC35316 ..

Align seg 1/1 to: AAC35316 from: 1 to: 643

```

6 AlalysProSerThrGluAspLeuGlyAspLys.....GlnGlyG1 20
:::||||| |||::: ||| ||| ||| ||| ||| ||| ||| |||
148 AGTAGATGCTGCTGCTCCGAGAGACAGACAGAGCTGACCAAGAGAC 197
20 utrylleuLysValIleGlyGlnAspSerSerGluIleHisPheL 37
:::|||||:||||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

198 TCACATCATCTCAAGTCAAGGACAGATGGTAATGAAGCTCTTTTA 247
37 ysValysMetThrThrHisLeuLysLeuLysGluSerTyrGln 53
:::||||| ::::::::::| ||| ||| ||| ||| ||| ||| |||
248 GGATCAAGACAGACACTCAGCTCAAAAACATGATGAATGCTTCTGAC 297
54 ArgGlnGlyValPromeTAsnSerLeuArgPheLeuPheGluGlnAr 70
|||||:||||| ::::::::::| ||| ||| ||| ||| ||| |||
298 GCTCATCTGTGATTTTCACATTCATTCGCTTCTTGTTCATGCTGCTG 347
70 gtleAlaSPAsnHisThrProLysGluLeuGlyMetGluGluAspY 87
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
348 TCTTCTGCTGCCAGACAGACATCCAGATGATGCTTGAATGAAGATG 397
87 allleGluValTyrGlnGluGlnThrGlyGly 97
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
398 AGATCGATCGAATGCTTCATCAGACTGCTGCT 429

seq_name: /SIDSL/gcdata/geneseq/geneseqn/NA2000.DAT:AAC48770
seq_documentation_block:
ID AAC48770 standard; DNA; 639 BP.
XX
AC AAC48770;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 58702.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132863.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.

```

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140623.
PR 29-JUN-1999; 99US-0140623.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142927.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0155658.
PR 29-SEP-1999; 99US-0155656.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157863.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 221.00 Length: 85
Ratio: 3.250 Gaps: 1
Percent Similarity: 80.000 Percent Identity: 51.765

alignment_block:
US-09-484-964-2 x AAC48770 ..

Align seg 1/1 to: AAC48770 from: 1 to: 639

```

15 Asplyslys.....GlnGlyGluTyrIleLysLeuLysValIleGlyG1 29
|||||.....|.....|.....|.....|.....|.....|.....|
172 GACAAAGAACCCGTGACCAAGAGCTCACATCATCTCAAGATCAAGGACAA 221
29 nAspSerSerGluIleHisPheLysValLysMetThrThrHisLeuLysL 46
|||||.....|.....|.....|.....|.....|.....|.....|
222 GGATGCTAATGAAAGCTCTTCTTAGGATCAAGAAAGCACTGACGTCACAAA 271
46 yslLeuLysGluSerTyrCysGlnArgIleAlaAspAsnHisThrProLysG1 62
|||||.....|.....|.....|.....|.....|.....|.....|
272 AACGTGATGAATGCTTACTGATGACCTGAGTCTGATTCATCACTCAAT 321
63 ArgPheLeuPheGluGluGlnArgIleAlaAspAsnHisThrProLysG1 79
|||||.....|.....|.....|.....|.....|.....|.....|
322 GCTTCTGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 371
79 uLeuGlyMetGluGluGlnAspValIleGluValTyrGlnGlnGluGlnT 96
|||||.....|.....|.....|.....|.....|.....|.....|
372 GCTTGAATGAAAGATGAGATGAGATGATGATGATGATGATGATGATGATG 421
96 IyGly 97
|||||
422 GTGCT 426

```

seq_name: /SIDS1/gcdata/geneseq/geneseqn/NA2000.DAT:AAA31186

seq_documentation_block:
ID AAA31186 standard; DNA: 351 BP.

```

XX AC AAA31186;
XX DT 05-JUL-2000 (first entry)
XX DE Plant microsatellite marker #147.
XX KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
XX variety identification; genetic variability evaluation; primer; ss.
XX OS EucaLyptus grandis.
XX PN W09967421-A1.
XX PD 29-DEC-1999.
XX PE 25-JUN-1999; 99WO-NZ00092.
XX PR 25-JUN-1998; 98US-0105307.
XX PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX PI Havukkalampi, Boksberg LN, Glenn M;
XX DR WPI; 2000-116958/10.
XX PT New plant microsatellite markers and associated flanking species for
XX the detection of polymorphic genetic markers -
XX Claim 1; Page 119; 392pp; English.

```

XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences AAA32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.

XX Sequence 351 BP: 97 A; 69 C; 112 G; 73 T; 0 other;

alignment_scores:
Quality: 218.50 Length: 87
Ratio: 3.213 Gaps: 1
Percent Similarity: 78.161 Percent Identity: 48.276

alignment_block:
US-09-484-964-2 x AAA31186 ..

Align seg 1/1 to: AAA31186 from: 1 to: 351

```

15 AsplyslysGlnGlyGlu.....TyrIleLysLeuLysValIleG1 28
|||||.....|.....|.....|.....|.....|.....|.....|
78 GACAAAGAACCCCGGATGACGCCGCCACATCAATCTGAAGTCAAGG 127
28 yGlnAspSerSerGluIleHisPheLysValLysMetThrThrHisLeuL 45
|||||.....|.....|.....|.....|.....|.....|.....|
128 CCAGATGCTATGAGTCTTTTATGATCAACAGCAAGCACCAGTTGA 177
45 yslLeuLysGluSerTyrCysGlnArgIleAlaAspAsnHisThrProLys 61
|||||.....|.....|.....|.....|.....|.....|.....|
178 GGAACCTCATGACCGCTACTGATGAGCAATCTGTGAGTAAATCTC 227
62 LeuArgPheLeuPheGluGluGlnArgIleAlaAspAsnHisThrProLys 78
|||||.....|.....|.....|.....|.....|.....|.....|
228 ATGCTCTTCTTATTTGATGAGCTGACTCTGCTGAGACAGACCTCTGA 277
78 sGluLeuGlyMetGluGluGlnAspValIleGluValTyrGlnGlnGlnT 95
|||||.....|.....|.....|.....|.....|.....|.....|
278 TGAGCTCGAGATGAGATGATGATGAATGACGACGATGCTCACAAA 327
95 hGlyGlyHis 98
|||||
328 CTGGGGGATAC 338

```

seq_name: /SIDS1/gcdata/geneseq/geneseqn/NA2000.DAT:AAC46716

seq_documentation_block:
ID AAC46716 standard; DNA: 505 BP.

```

XX AC AAC46716;
XX DT 18-OCT-2000 (first entry)
XX DE Zea mays DNA fragment SEQ ID NO: 51143.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic;
XX pathway; promoter; termination sequence; corn; ss.
XX OS Zea mays subsp. mays.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.

```

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0122548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 03-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138099.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

APPLICANT: Hillman, Jennifer L.
 TITLE OF INVENTION: HUMAN SMT3-LIKE PROTEIN
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/853,974
 FILING DATE: Filed Herewith
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0289 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 473 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: LUNGFET03
 CLONE: 2361410
 US-08-853-974-2

alignment_scores:
 Quality: 217.00 Length: 98
 Ratio: 2.893 Gaps: 2
 Percent Similarity: 76.531 Percent Identity: 44.898

alignment block:
US-09-484-964-2 x US-08-853-974-2

Align seg 1/1 to: US-08-853-974-2 from: 1 to: 473

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysly 17
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 ATGTCCGAGGAGAACCCCAAGAG.....GGTGTCAAGAC 118
17 sGlnGlyIuTyrlLysLeuLysValIleGlyLnsPserSergIui 34
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 AGAGATGACCAATCACTGAAGGTGGCCGGGCGAGGACGGTCCTCGTGG 168
34 leHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 TGCAGTTCAGATCAAGCAGCAGCGCGGTGAGCAGCAAGCTGATGAAGGCC 218
51 TyrcysGlnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheG1 67
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 TACGCCAGAGAGCGAGGCTTGTCAATATGAGCAGATCAGATTCAAGTTGCA 268
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
269 CGGGCAGCCCAATCAATGAAGTACACTCCAGACACAGCTGAGATGAGAGG 318
84 lUGluAspValIleGluValTyT...GlnGluGlnThrGlyGly 97
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

319 ACGAGACACCATGACGCTGTTTCCAAAGCAGACAGCGGAGGT 362
 seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:US-09-172-988-2
 seq_documentation block:
 ; Sequence 2, Application US/09172988
 ; Patent No. 6136935
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Shah, Purvi
 ; TITLE OF INVENTION: HUMAN SMT3-LIKE PROTEIN
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/172,988
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/853,974
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0289 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 473 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: LUNGFET03
 ; CLONE: 2361410
 ; US-09-172-988-2

alignment_scores:
 Quality: 217.00 Length: 98
 Ratio: 2.893 Gaps: 2
 Percent Similarity: 76.531 Percent Identity: 44.898

alignment block:
US-09-484-964-2 x US-09-172-988-2

Align seg 1/1 to: US-09-172-988-2 from: 1 to: 473

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysly 17
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 ATGTCCGAGGAGAACCCCAAGAG.....GGTGTCAAGAC 118
17 sGlnGlyIuTyrlLysLeuLysValIleGlyLnsPserSergIui 34
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 AGAGATGACCAATCACTGAAGGTGGCCGGGCGAGGACGGTCCTCGTGG 168
34 leHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 TGCAGTTCAGATCAAGCAGCAGCGCGGTGAGCAGCAAGCTGATGAAGGCC 218
51 TyrcysGlnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheG1 67
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

15 .....AsplylsylgInglYtIlyleLysl 24
990 GGTCGTAGCAACGCTTCCAGAAGGCCAACAGACATCCCAGAGACTCGGC 1039
24 eulyValIlleGlYGlnaspsrserGu...lleHisphelyVallys 39
||:::|| |::::::::::| ::::::::::::::::::::|
1040 TCCGGTGCGAAGGAAGAAACAACAGATGTTGGAATCCTCACGTCT 1089
40 MetThrThlslseulslysLeulsygsIsertyrcYsglnayggngl 56
::: |||||| |:: ||:: ||:: :||
1090 CCGATTCTTCCTTAAGTCTCATGTCACACTATGAGGAAGCCATGG 1139
56 yalPrometAnSerLeuarArgPheuePhelugluGlnatrgllleaiaa 73
::: |:|:::| |::|::|::|::|::|::|::|::|::|::|
1140 ACCTGTGACACAACTCCTCTCTCTTTGATGGAGCAAAGCTTTCAG 1189
73 spasnHsthrProLySGLulenclYmetcLuIgLnuspaVallleclu 89
:|:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1190 GCAAGGAGCTGCCAGCTGATCTGGCGCTGGAATCCGAGATCTCATCGMA 1239
90 Valtyr 91
||:::|
1240 GTCTGG 1245

seq_name: /cnrl_6/prodcta/2/ina/6A_COMB.seq:US-09-192-611-1
seq_documentation_block:
Sequence 1, Application US/09192611
Patent No. 6090561
GENERAL INFORMATION:
APPLICANT: Glmcher, Laurie H.
APPlicant: Dodge, Martin R.
TITLE OF INVENTION: NF-AF-INTERACTING PROTEIN NIP4S AND METHODS
OF USE THEREFOR
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/192,611
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/755,584
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Giulio A. Decortti, JT.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: HUT-026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 13..1248
US-09-192-611-1
```

us-09-484-964-2.rn1

REGISTRATION NUMBER: 39,317

.. 5510474-1 X 2 X 904-2 05-484

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Thu Aug 16 08:48:54 2001

us-09-484-964-2.rni

Page 5

```
13 LeuGlyAspLysLysGlnGlyGluTyrIleLysLeuLysValIleGly1 29
|||||
2652 CTGCGTCTGCGCTGAGGAGGAGCATGAGATCTTCTGGA...GACC 2698
29 nAspSerSerGluIleHis.PheLysValLysMetThrThrHisLeuLys 45
|||
2699 TGACGGGCAAGACTATCACCCTCGAGGTGAGCTTCACAGACCATCGAC 2748
46 LysLeuLysGluSerTyrCysGlnArgGlnGlyValProMetAsnSerle 62
|||||
2749 AATGTCAAGGCCAAGATCCAGACAGAGGCGATCCACCGGACCA 2798
62 uArgPheLeuPheGlnGlyGlnArgIleAlaAspAsnHisThrProLysG 79
|||||
2799 GCGTTGTATCTTCCCTGCAAGCAGCTGAGAGATGGCCGACCTTGGCG 2848
79 LuLeuGlyMetGluGluGluAspValIleGluValTyrGlnGlnGluThr 95
:::
2849 ATTACAACATCCAGAGAGAGACCTCCAGCTGCTGCTCCTCAGG 2898
96 GlyGly 97
|||||
2899 GGTGGT 2904

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-462-092-1

seq_documentation block:
; Sequence 1, Application US/08462092
; Patent No. 5614399
;
; GENERAL INFORMATION:
; APPLICANT: Peter H. Quail
; APPLICANT: Alan H. Christensen
; APPLICANT: Howard P. Hershey
; APPLICANT: Robert A. Sharrock
; APPLICANT: Thomas D. Sullivan
; TITLE OF INVENTION: PLANT UBIQUITIN PROMOTER SYSTEM
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,092
; FILING DATE: 5-JUNE-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: 08/462,092
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-375-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3840 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1993..3591
;
; US-08-462-092-1
```

```
alignment_scores:
  quality: 78.50      length: 86
  ratio: 1.402      gaps: 2
  percent similarity: 65.116      percent identity: 22.093

alignment block:
US-09-484-964-2 x US-08-462-092-1 ..
Align seg 1/1 to: US-08-462-092-1 from: 1 to: 3840

13 LeuGlyAspLysLysGlnGlyGluTyrIleLysLeuLysValIleGly1 29
|||||
2424 CTGCGTCTGCGCTGAGGAGGAGCATGAGATCTTCTGGA...GACC 2470
29 nAspSerSerGluIleHis.PheLysValLysMetThrThrHisLeuLys 45
|||
2471 TGACGGGCAAGACTATCACCCTCGAGGTGAGCTTCACAGACCATCGAC 2520
46 LysLeuLysGluSerTyrCysGlnArgGlnGlyValProMetAsnSerle 62
|||||
2521 AATGTCAAGGCCAAGATCCAGACAGAGGCGATCCACCGGACCA 2570
62 uArgPheLeuPheGlnGlyGlnArgIleAlaAspAsnHisThrProLysG 79
|||||
2571 GCGGTCTATCTTCCCTGCAAGCAGCTGAGAGATGGCCGACCTTGGCG 2620
79 LuLeuGlyMetGluGluGluAspValIleGluValTyrGlnGlnGluThr 95
:::
2621 ACTACAACATCCAGAGAGAGACCTCCAGCTGCTGCTCCTCAGG 2670
96 GlyGly 97
|||||
2671 GGAGGC 2676

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-746-822-1

seq_documentation block:
; Sequence 1, Application US/08746822
; Patent No. 6020190
;
; GENERAL INFORMATION:
; APPLICANT: Peter H. Quail
; APPLICANT: Alan H. Christensen
; APPLICANT: Howard P. Hershey
; APPLICANT: Robert A. Sharrock
; APPLICANT: Thomas D. Sullivan
; TITLE OF INVENTION: PLANT UBIQUITIN PROMOTER SYSTEM
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,822
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,092
; FILING DATE: 5-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: 08/462,092
; TELECOMMUNICATION INFORMATION:
```


Page 7

```

alignment_scores:
  Quality: 77.50      Length: 112
  Ratio: 1.174      Gaps: 2
  Percent Similarity: 58.929      Percent Identity: 19.643

alignment_block:
  US-09-484-964 x US-08-284-784-42  ..

Align seg 1/1 to: US-08-284-784-42 from: 1 to: 915

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
   :::::  :::::|||||:  :::::  |||
367 ATCAGCACCTTCACGAGAAACATGCCGCTCTTCTCGGAGAGGACCA 416
   :::::  :::::|||||:  :::::  |||

17 SGInGluGlyIrrIle.....LysLeuLysValIleGlyGlnAsp 31
   |  ||:::|  :::::  :::::  :::::  |||
417 AGCGCGCCAGGATATTAAGTACTGACTTCACATGCAATTTGTCTTCCGACC 466
   |  ||:::|  :::::  :::::  :::::  |||

31 eRserGluIleHisPheLysValLysMetIrrThr..... 42
   ::  :::::  |||  |||  |||
467 GCGGTGGCATGCAGATTTTCGTCAAGCTTTGACCGGTAAACCATACACA 516
   ::  :::::  |||  |||  |||

43 .....HisLeuLysLysLeuLysGluSerIrrCy 52
   ::  :::::  |||  |||  |||

517 TTGGAAAGTGAATCTTCGCATACCACTGCACACAGTTAAGTGAATAATCA 566

52 sGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheGluGly 69
   :::::  |||  :::::  :::::  |||  |||
567 AGCAACAGAGAGATCCCTCCAGATCAACAAAGATGTGATCTTTCCSGGTA 616

69 lnaRgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlnGlu 85
   :::::  |||  :::::  :::::  |||  |||
617 AGCGAGTGAAGACGGTGAACCGCTCTGATTTACATCAATTCACAGAGAG 666

```

17 sg]ng]vg]utvr]le.....L

```

seq_name: /cgn2_6/ptodata/2/lna/5B.COMB.seq:US-08-854-811-42
seq_documentation_block:
Sequence 42, Application US/08854811
Patent No. 5914254
GENERAL INFORMATION:
APPLICANT: Mascarenhas, Desmond
APPLICANT: Zhang, Yang
APPLICANT: Olson, Pamela S.
APPLICANT: Olsen, David R.
APPLICANT: Cohen, Pedro A.
TITLE OF INVENTION: EXPRESSION OF FUSION POLYPEPTIDES
TITLE OF INVENTION: TRANSPORTED OUT OF THE CYTOPLASM WITHOUT LEADER
TITLE OF INVENTION: SEQUENCES
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSER: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,811
FILING DATE: 12-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284,784
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/100,744
FILING DATE: 02-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bufiger, Nicholas S
REGISTRATION NUMBER: 39,124
REFERENCE/DOCKET NUMBER: 22095-20275.21
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-854-811-42

alignment_scores:
Quality: 77.50 Length: 112
Ratio: 1.174 Gaps: 2
Percent Similarity: 58.929 Percent Identity: 19.643

alignment_block:
US-09-484-964-2 x US-08-854-811-42 ..

Align seg 1/1 to: US-08-854-811-42 from: 1 to: 915

1 MetSerAspGlnGlnAlaTyrProSerThrGlnSpluGlnLysAspLysI 17
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
367 ATCAACACCTCTCAAGCAAAACATGCCGCTTCTCGGAGAGACCA 416
17 scGlnuYsGlnuYrIle.....LysLeuYsValIleGlyGlnAsps 31

```

alignment_scores:

FILING DATE: 11-08-93
 ATTORNEY/AGENT INFORMATION:
 NAME: WASSERMAN, FRAN S.
 REGISTRATION NUMBER: 34,273

PRIOR APPLICATION DATA:

[illegible]

Thu Aug 16 08:48:54 2001

us-09-484-964-2.rni

Page 10

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd

OM protein - protein search, using sw model

Run on: August 15, 2001, 16:37:34 ; Search time 73.79 Seconds
(without alignments)
104.264 Million cell updates/sec

Title: US-09-484-964-2
Page: 535

Sequence: 1 MSDQEAKPSTEDLGDKQGE.....MEEDVIEVYQEQTGCHSTV 101

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```
Database :
1: pir1:
2: pir2:
3: pir3:
4: pir4:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	280.5	53.4	91	2	JG5582	SMR3 protein - Cae
2	244.5	46.6	117	2	T40313	ubiquitin-like mod
3	233.5	44.5	90	2	T43537	ubiquitin-like pro
4	232	44.2	95	2	JC4760	SMR3 protein - hum
5	223.5	42.6	100	2	T04807	SMR3 protein homol
6	222	42.3	100	2	T04102	smr3 protein - ric
7	218.5	41.6	101	2	S63959	SMR3 protein - yea
8	208	39.6	115	2	T09529	ubiquitin-like pro
9	147	28.0	380	2	T00792	hypothetical prote
10	98	18.7	154	2	S5242	polyubiquitin 2 -
11	98	18.7	156	2	C86439	protein T19E23.13
12	96	18.3	78	2	G86254	hypothetical prote
13	93.5	17.8	456	2	S61158	hypothetical prote
14	88.5	16.9	655	1	S50096	probable splicing
15	85.5	16.3	144	2	A96580	hypothetical prote
16	85.5	16.3	406	2	T39786	hypothetical prote
17	82.5	15.7	551	2	C84549	probable ubiquitin
18	81	15.4	77	1	U0NVAC	ubiquitin - Antogr
19	81	15.4	77	2	C72854	viral ubiquitin
20	81	15.4	77	2	T41781	ubiquitin orf35 -
21	81	15.4	264	2	S62809	ubiquitin precursor
22	81	15.4	381	2	A27806	polyubiquitin 5 (C
23	80	15.2	128	1	U0DOR	ubiquitin / riboso
24	80	15.2	150	2	T30390	probable ubiquitin
25	80	15.2	154	1	U0DOR7	ubiquitin / riboso
26	80	15.2	228	2	D34080	ubiquitin 18 - sil
27	80	15.2	229	1	B27805	ubiquitin (clone 1
28	80	15.2	380	2	C34080	polyubiquitin 5 (C
29	80	15.2	380	2	B34080	polyubiquitin 5 (C

30	80	15.2	442	2	B71083	hypothetical prote
31	80	15.2	532	2	A34080	polyubiquitin 7 (c
32	79	15.0	76	2	S31653	ubiquitin - Giardi
33	79	15.0	78	2	T39665	ubiquitin-like pro
34	79	15.0	80	2	T28305	ORF MSV144 probabl
35	78.5	15.0	138	2	S62740	ubiquitin precursor
36	76.5	14.6	388	2	T24432	hypothetical prote
37	76	14.5	81	2	JN0710	ubiquitin-like pro
38	76	14.5	188	2	S45304	ubiquitin / riboso
39	76	14.5	177	2	T24466	hypothetical prote
40	76	14.5	274	2	G84602	hypothetical prote
41	75.5	14.4	379	2	S55245	polyubiquitin 5 -
42	75	14.3	129	2	JC5226	ubiquitin / riboso
43	75	14.3	133	2	A49768	polyubiquitin - see
44	75	14.3	149	1	UNOCR	ubiquitin / riboso
45	75	14.3	154	2	T46664	ubiquitin/527a fus

ALIGNMENTS

```

RESULT 1
JC5582
SMT3 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 05-Nov-1999
C:Accession: JC5582; r32928
R:Choudhury, B.K.; Li, S.S.L.
Biochem. Biophys. Res. Commun. 234, 788-791, 1997
A:Title: Identification and characterization of the SMT3 cDNA and gene from nematode
A:Reference number: JC5582; MUID:97318858
A:Accession: JC5582
A:Molecule type: DNA
A:Residues: 1-91 <CHO>
A:Cross-references: GB:U94830; NID:g2341094; PIDN:AAB67608.1; PID:g2341095
R:Wamsley, P.; Kramer, J.
Submitted to the EMBL Data Library, January 1998
A:Description: The sequence of C. elegans cosmid K12C11.
A:Reference number: 221248
A:Accession: r32928
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-91 <NMC>
A:Cross-references: EMBL:AF043701; PIDN:AAB97577.1; GSPDB:GND0019; CESP:K12C11.2
A:Experimental source: Strain Bristol N2; clone K12C11
C:Comment: This protein is a suppressor of Mif2 mutation in a centromere protein.
A:Genetics:
A:Gene: smt3; CESP:K12C11.2
A:Map position: 1
A:Introns: 22/3; 56/2
C:Superfamily: yeast SMT3 protein

Query Match 53.4%; Score 280.5; DB 2; Length 91;
Best Local Similarity 58.8%; Pred. No. 9,4e-21;
Matches 57; Conservative 15; Mismatches 18; Indels 7; Gaps 2;

OY 1 MSDOEAKPSTEDJGDKKQGYIKLVYIGDSSSTHKKVYKMTHLKKIKESYCGRCGYVMN 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MADDAAO-----AGD--NAEYIKIKVYGQDSNVEHRYKGYSTMAKLKRSYADRTGYAVN 53
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 61 SLRFLFGORLADNHPPKELGMEEDVIEYQEOGTC 97
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 54 SLRFLFGRRINDDTPKLTLEMEDDVIEYQEOGLG 90
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
T40313
ubiquitin-like modifier protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Feb-2000
C:Accession: T40313; T43416
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lelaure, V.; Galbert, F.

```


1000 #text change 20-Jun-2000

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999
C:Accession: T09529
R:Labrador, E.
submitted to the EMBL Data Library October 1997
A:Description: Isolation and characterization of a cDNA sequence CansMT3 from *Cicer a*
A:Description number: 216717
A:Reference number: T09529
A:Accession: T09529
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-115 <LAB>
A:Cross-references: EMBL:AF001901
A:Experimental source: cultivar Castellana; 5 d old seedlings; etiolated epicotyls
C:Superfamily: Yeast SMT3 protein

[illegible]

RESULT 9
T00792
hypothetical protein At2g32760 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F24L7.10
C:Species: Arabidopsis thaliana (mouse ear cress)
C:Date: 12-Feb-1999 #sequence,revision 12 Feb-1999 #text,change 16-Feb-2001
C:Accession: T00792; B64373
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
Submitted to the EMBL Data library, February 1998
A:Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.
A:Reference number: Z14204
A:Accession: T00792
A:Status: translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-380 <R0U>
A:Cross-references: EMBL:AC003974; NID:g2914688; PID:g2914697
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, E.; Tallon,
N.; Koo, H.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Andrew, D. 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: B64420; MUID:20083487
A:Accession: B64737
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-380 <S7O>
A:Cross-references: GB:AE002093; NID:g2914697; PID:AA04487.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g32760; F24L7.10
A:Map position: 2
A:Introns: 55/3; 116/1; 161/3; 237/3; 287/3; 357/3

Query Match	Best local	Similarity	Score	DB 2:	Length	380:
Matches	30:	Conservative	147:	15:	Indels	Gaps
QY	22	IKKKYIGDSSHHFKVKMTHTLKKIKESYCOGQVPMNSIRFLFCQKIDNHPKELG	28.0% 39.5%	DB 2: Pred. No. 5.6e-07;	Length 31:	380: 0;
Db	300	VTLKVNQOGADLYIKIGTNAHKLMSATCKRNLDSVRYVYNGRIKARQTPALDH	30:	Conservative	15:	Indels
QY	82	MEEDVIEYVEQETGG	97			
Db	360	MEEDELCYMELGGG	375			

RESULT 10

S55242

polyubiquitin 2 - Arabidopsis thaliana

N:Alternate names: ubiquitin-like protein

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 11-Apr-1997

R:Callis, J.; Carpenter, T.; Sun, C.W.; Vierstra, R.D.

A:Title: Structure and evolution of genes encoding polyubiquitin and ubiquitin-like proteins

A:Reference number: S55242; MUID:95229071

A:Accession: S55242

A:Molecule type: DNA

A:Cross-references: EMBL:L05364

A:Experimental source: ecotype Columbia

C:Genetics:

A:Introns: 31/3; 118/2

C:Superfamily: polyubiquitin 2; ubiquitin homology

F:1-76/Product: ubiquitin #status Predicted <UB1>

F:77-152/Product: ubiquitin #status Predicted <UB2>

F:77-152/Product: ubiquitin homology <UBH2>

Query Match

Best Local Similarity 18.7%; Score 98; DB 2; Length 154;

Matches 21; Conservative 21; Mismatches 35; Indels 2; Gaps 1;

OY 19 GEYIKLVIGDSSSEIHFKVMTHTLKKLESYCOGQVPMNSLRFLFEGQRIADNHPTK 78

Db 76 GYMKVKTLL--TKREIDIEPTDITDRIKEREKEGIPVQOQLTYAGKQADADTKAK 133

OY 79 ELGNEEDVIEVYQEQGTG 97

Db 134 DYALGGSVHLVLAALRG 152

RESULT 11

C86439

Protein T19E23.13 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

R:Heide, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Hutzar, L.

A:Title: Structure and evolution of genes encoding polyubiquitin and ubiquitin-like proteins

A:Reference number: S55242; MUID:95229071

A:Accession: S55242

A:Molecule type: DNA

A:Cross-references: EMBL:L05364

A:Experimental source: ecotype Columbia

C:Genetics:

A:Introns: 31/3; 118/2

C:Superfamily: polyubiquitin 2; ubiquitin homology

F:1-76/Product: ubiquitin #status Predicted <UB1>

F:77-152/Product: ubiquitin #status Predicted <UB2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

OY

19

GEYIKLVIGDSSSEIHFKVMTHTLKKLESYCOGQVPMNSLRFLFEGQRIADNHPTK 78

Db

76

GYMKVKTLL--TKREIDIEPTDITDRIKEREKEGIPVQOQLTYAGKQADADTKAK 133

OY

79

ELGNEEDVIEVYQEQGTG 97

Db

134

DYALGGSVHLVLAALRG 152

RESULT 12

G86254

hypothetical protein [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

R:Heide, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Hutzar, L.

A:Title: Structure and evolution of genes encoding polyubiquitin and ubiquitin-like proteins

A:Reference number: S55242; MUID:95229071

A:Accession: S55242

A:Molecule type: DNA

A:Cross-references: EMBL:L05364

A:Experimental source: ecotype Columbia

C:Genetics:

A:Introns: 31/3; 118/2

C:Superfamily: polyubiquitin 2; ubiquitin homology

F:1-76/Product: ubiquitin #status Predicted <UB1>

F:77-152/Product: ubiquitin #status Predicted <UB2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

F:77-152/Product: ubiquitin homology <UBH2>

Db 382 EYARIALMGQDNKKIYVHRSTPFSKIAEYRIQKQLPQKTRVKLLFDHDELDNMCEIA 441
 QY 79 ELGMEEDVIEV 90
 Db 442 DQMEDEDMVDV 453

RESULT 14

probable splicing factor Ccprp21 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: S50096; T33806

R:Spikes, D.A.; Kramer, J.; Bingham, P.M.; van Doren, K.

Nucleic Acids Res. 22, 4510-4519, 1994

A:Title: SWAP pre-mRNA splicing regulators are a novel, ancient protein family sharing a

A:Reference number: S50095; MUID:95061415

A:Accession: S50096

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-655 <SP1>

A:Cross-references: EMBL:U09415; NID:9498845; PIDD:AAA2164.1; PID:9498846

R:Latreille, P.; Wamsley, P.

submitted to the EMBL Data Library, November 1998

A:Description: The sequence of C. elegans cosmid w07E6.

A:Reference number: 221414

A:Accession: T33806

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-655 <LAT>

A:Cross-references: EMBL:AF106576; PIDD:AAC78179.1; GSPDB:GN00020; CESP:w07E6.4

A:Experimental source: strain Bristol N2; clone w07E6

C:Genetics:

A:Gene: CESP:w07E6.4

A:Map position: 2

A:Introns: 36/1; 132/2; 185/3; 309/1; 402/2; 464/1; 532/3; 598/3

C:Superfamily: Caenorhabditis elegans probable splicing factor Ccprp21

Query Match 16.9%; Score 88.5; DB 1; Length 655;

Best Local Similarity 24.5%; Pred. No. 0.55;

Matches 26; Conservative 20; Mismatches 43; Indels 17; Gaps 3;

QY 7 KPSTEDLGD-----KQGEYIKLV-----IGDSSIHFRKVTTLKKIKESY 51

Db 549 RPTED--DLIPEDMLKRVNGAISLNVHLPAPEHGMDGISVFTIQTAPMSELKQOI 606

QY 52 CORQGVPMNSLRFEGORIDNHFPKELGMEEDVIEVQEOGTG 97

Db 607 QDRYGMFVGKQKIMSDGLFVKDNMSAFYNLADRTAITYIQVKERGS 652

RESULT 15

hypothetical protein F1511.4 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: A96580

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Raul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maltl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: A96580

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-144 <STO>

A:Cross-references: GB:AE005173; NID:94587536; PIDD:AAD25767.1; GSPDB:GN00141

C:Genetics:

A:Gene: F1511.4

A:Map position: 1

Query Match

Best Local Similarity 16.3%; Score 85.5; DB 2; Length 144;

Matches 23; Conservative 27; Mismatches 27; Indels 41; Gaps 3;

QY 18 QGEYIKLVIGQDSS-----EIH-----FKV 38

Db 28 EGKRIKLEV--EDSSNTIDKIHGPTRELVDLSPPTDGTAIMMIFIKTLTGRTNYEV 85

QY 39 KMTTLKRLKLSYCORQGVPMNSLRFEGORIDNHFPKELGMEEDVIEVQEOGTG 96

Db 86 KGSPTIRRLKAKHKEKEGIPVEOORLIFQGRVLDSDSKAISPYNIKHESTLHITLHOCG 143

Search completed: August 15, 2001, 17:16:54

Job time: 2360 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: August 15, 2001, 17:11:49 ; Search time 57.57 Seconds
(without alignments)
60.097 Million cell updates/sec

Title: US-09-484-964-2

Perfect score: 525
Sequence: 1 MSDQKAPSTEDIGDKKGE.....MEEDVIEYVQPTGSHSTV 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 segs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt.39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	99.4	101	1	SM33_HUMAN
2	280.5	53.4	91	1	SM33_CAEEL
3	244.5	46.6	117	1	SM33_SCHPO
4	239.5	45.6	103	1	SM33_HUMAN
5	232	44.2	95	1	SM32_HUMAN
6	223.5	42.6	104	1	SM33_ARATH
7	222	42.3	100	1	SMT3_ORYSA
8	218.5	41.6	101	1	SMT3_YEAST
9	81	15.4	77	1	UBIL_NPYAC
10	80	15.2	76	1	UBIL_DICDI
11	76	14.5	76	1	UBIQ_ACACA
12	76	14.5	81	1	NED8_HUMAN
13	75	14.5	81	1	NED8_MOUSE
14	75	14.3	76	1	UBIQ_NEUCR
15	75	14.3	76	1	UBIQ_STRPU
16	74	14.1	76	1	UBIQ_ACEEL
17	74	14.1	76	1	UBIQ_AGLEL
18	74	14.1	76	1	UBIQ_CAEEL
19	74	14.1	76	1	UBIQ_PHYTN
20	73.5	13.9	594	1	YJG9_YEAST
21	73	13.9	76	1	UBIQ_CHLRE
22	73	13.9	76	1	UBIQ_HUMAN
23	73	13.9	76	1	UBIQ_TRYCR
24	73	13.9	76	1	UBIQ_TRYCR
25	72.5	13.8	476	1	UBIQ_YEAST
26	72	13.7	76	1	UBIQ_COPCO
27	72	13.7	76	1	UBIQ_TETPY
28	72	13.7	93	1	UBIL_NPYOP
29	71	13.5	76	1	UBIQ_EIMBO
30	71	13.5	76	1	UBIQ_SOYHN
31	70.5	13.4	838	1	OS94_MOUSE
32	70.5	13.4	523	1	KPT2_HUMAN
33	70.5	13.4	523	1	KPT2_RAT

34	70	13.3	76	1	UBIQ_EUPBU	P23324 euplotes eu
35	70	13.3	153	1	UCRP_BOVIN	002741 bos taurus
36	69	13.1	76	1	UBIQ_LETMA	005550 leishmania
37	69	13.1	76	1	UBIQ_TRYBA	P15174 trypanosoma
38	69	13.1	352	1	CLCE_PSEAE	009612 pseudomonas
39	69	13.1	476	1	YSS4_CAEEL	009663 caenorhabd
40	68.5	13.0	380	1	KPT3_HUMAN	007002 homo sapien
41	68	13.0	153	1	MYG_ZALCA	P02161 zalophus ca
42	68	13.0	352	1	CLCE_PSEAB	030847 pseudomonas
43	68	13.0	355	1	BAG1_MOUSE	060739 mus musculu
44	67	12.8	373	1	DSK2_YEAST	P48510 saccharomyc
45	67	12.8	868	1	NRG2_RAT	035569 rattus norv

ALIGNMENTS

RESULT	1	STANDARD	PRT	101 AA.
SM33_HUMAN				
ID	SM33_HUMAN			
AC	Q93068; P55856;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	UBIQUITIN-LIKE PROTEIN SMT3C (UBIQUITIN-HOMOLOGY DOMAIN PROTEIN PIC1)			
DE	(UBIQUITIN-LIKE PROTEIN UBL1) (UBIQUITIN-RELATED PROTEIN SUMO-1)			
DE	(GAP MODIFYING PROTEIN 1) (GMP1) (SENTRIN).			
GN	UBL1 OR SMT33 OR SMT3C.			
OS	Homo sapiens (Human), and Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid=9606, 10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Human; TISSUE=Brain;			
RA	MEDLINE=97237059; PubMed=9119407;			
RA	Lapenta V., Chitrazzi P., van der Spek P.J., Pizutti A.,			
RA	Hanaoka F., Brahe C.;			
RT	"SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to			
RT	chromosome 21qter and defines a novel gene family.";			
RT	Genomics 40:362-367(1997).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	SPECIES=Human; TISSUE=Placenta;			
RC	MEDLINE=96400311; PubMed=8806687;			
RA	Boddy M.N., Howe K., Etkin L.D., Solomon E., Freemont P.S.;			
RT	"Ptc 1, a novel ubiquitin-like protein which interacts with the PML			
RT	component of a multiprotein complex that is disrupted in acute			
RT	promyelocytic leukaemia.";			
RT	Oncogene 13:971-982(1996).			
RL	[3]			
RN	SEQUENCE FROM N.A.			
RP	SPECIES=Human;			
RC	MEDLINE=96411684; PubMed=8812453;			
RA	Shen Z., Pardington-Puttyman P.E., Comeaux J.C., Moyzis R.K.,			
RT	Chen D.J.;			
RT	"UBL1, a human ubiquitin-like protein associating with human			
RT	RAD51/RAD52 proteins.";			
RT	Genomics 36:271-279(1996).			
RL	[4]			
RN	SEQUENCE FROM N.A.			
RP	SPECIES=Human;			
RC	MEDLINE=97148692; PubMed=9019411;			
RA	Mahajan R., Delphin C., Guan T., Gerace L., Melchior F.;			
RT	"A small ubiquitin-related polypeptide involved in targeting RANGAP1			
RT	to nuclear pore complex protein RanBP2.";			
RT	Cell 88:97-107(1997).			
RL	[5]			
RN	SEQUENCE FROM N.A.			
RP	SPECIES=Human;			
RC	MEDLINE=97133418; PubMed=8978815;			
RA	Matunis M.J., Coutavas E., Blobel G.;			
RT	"A novel ubiquitin-like modification modulates the partitioning of			

RT the Ran-GTPase-activating protein RANGAP1 between the cytosol and the
 RT nuclear pore complex."
 RT J. Cell Biol. 135:1457-1470(1996).
 RN
 RN
 RP SEQUENCE FROM N.A.
 RC SPECIES-Human; TISSUE-Placenta;
 RX MEDLINE-97064180; PubMed-8906799;
 RA Okura T., Gong L., Kamitani T., Wada T., Okura I., Wei C.F.,
 RA Chang H.M., Yeh E.T.H.;
 RT "Protection against Fas/Apo-1- and tumor necrosis factor-mediated
 RT cell death by a novel protein, sentrin";
 RL J. Immunol. 157:4277-4281(1996).
 RN
 RP SEQUENCE FROM N.A.
 RC SPECIES-Mouse; STRAIN-ICR;
 RX MEDLINE-98126440; PubMed-9465300;
 RA Howe K., Williamson J., Boddy M.N., Sheer D., Freemont P.S.,
 RA Solomon E.;
 RT "The ubiquitin-homology gene P1C1: characterization of mouse (P1C1)
 RT and human (UBI1) genes and pseudogenes.";
 RL Genomics 47:92-100(1998).
 RN
 RP STRUCTURE BY NMR.
 RC SPECIES-Human;
 RX MEDLINE-98319859; PubMed-9654451;
 RA Bayer P., Arndt A., Metzger S., Mahajan R., Melchior F., Jaenicke R.,
 RA Becker J.;
 RT "Structure determination of the small ubiquitin-related modifier
 RT SUMO-1";
 RL J. Mol. Biol. 280:275-286(1998).
 CC -1- FUNCTION: ASSOCIATES WITH RAD51/RAD52. INVOLVED IN TARGETING
 CC RANGAP1 TO THE NUCLEAR PORE COMPLEX PROTEIN RANBP2.
 CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X99586; CAA67898.1; -
 DR EMBL: U61397; AAB40388.1; -
 DR EMBL: U38784; AAC50733.1; -
 DR EMBL: U67122; AAC50996.1; -
 DR EMBL: U72722; AAB40390.1; -
 DR EMBL: U83117; AAB39999.1; -
 DR EMBL: AF033353; AAC39959.1; -
 DR PDB: 1A5R; 14-OCT-98.
 DR MIN: 601912; -
 DR MGD: MGI:1197010; DB11.
 DR InterPro: IPR000626; -
 DR PROSITE: PSS0053; UBIQUITIN_2; 1.
 DR 3D-structure.
 KW DOMAIN 20 101 UBIQUITIN-LIKE.
 FT SEQUENCE 101 AA; 11557 MW; 89BE97D2D054FB33 CRC64;
 SQ

Query Match 99.4%; Score 522; DB 1; Length 101;
 Best Local Similarity 99.0%; Pred. No. 2e-44;
 Matches 100; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDDAKRSTEDLDGKRGEGYIKLVIGODSSRHFVKMTTHLKKLKESSYQROGVPMN 60
 DB 1 MSDDAKRSTEDLDGKRGEGYIKLVIGODSSRHFVKMTTHLKKLKESSYQROGVPMN 60
 QY 61 SLRFLFGORLADNHTPKRELGMEEEDVIEVYQEGTGHSTV 101
 DB 61 SLRFLFGORLADNHTPKRELGMEEEDVIEVYQEGTGHSTV 101

RESULT 2

SMT3-CAEEL
 ID SMT3-CAEEL STANDARD; PRT; 91 AA.
 AC P55853;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 GN UBIQUITIN-LIKE PROTEIN SMT3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97237059; PubMed-9119407;
 RA Lapenta V., Chiuazzoli P., van der Spek P.J., Pizzuti A.,
 RA Hanoka F., Brahe C.;
 RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
 RT chromosome 21qter and defines a novel gene family.";
 RL Genomics 40:362-367(1997).
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97318858; PubMed-9175795;
 RA Choudhury B.K., Li S.S.;
 RT "Identification and characterization of the SMT3 cDNA and gene from
 RT nematode Caenorhabditis elegans.";
 RL Biochem. Biophys. Res. Commun. 234:788-791(1997).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Wamsley P., Kramer J.;
 RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X99600; CAA67914.1; -
 DR EMBL: U94830; AAB67608.1; -
 DR EMBL: AF043701; AAB97577.1; -
 DR WormPep: K12C11.2; CEI8056.
 DR InterPro: IPR000626; -
 DR PROSITE: PSS0053; UBIQUITIN_2; 1.
 DR DOMAIN 13 91 UBIQUITIN-LIKE.
 FT SEQUENCE 91 AA; 10222 MW; 0894E9B6F7B37P5 CRC64;
 SQ

Query Match 53.4%; Score 280.5; DB 1; Length 91;
 Best Local Similarity 58.8%; Pred. No. 5.9e-21;
 Matches 57; Conservative 15; Mismatches 18; Indels 7; Gaps 2;

QY 1 MSDDAKRSTEDLDGKRGEGYIKLVIGODSSRHFVKMTTHLKKLKESSYQROGVPMN 60
 DB 1 MADDAQ-----AGD--NAEYIKLVIGODSSRHFVKMTTHLKKLKESSYQROGVPMN 53
 QY 61 SLRFLFGORLADNHTPKRELGMEEEDVIEVYQEGTGHSTV 97
 DB 54 SLRFLFGORLADNHTPKRELGMEEEDVIEVYQEGTGHSTV 90

RESULT 3
 SMT3-SCHEP
 ID SMT3-SCHEP STANDARD; PRT; 117 AA.
 AC O13351; 074186;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 GN UBIQUITIN-LIKE PROTEIN SMT3/PWT3.
 DE

GN SMT3 OR PMT3 OR SPC365.06.
 OS Schizosaccharomyces pombe (Fission yeast).
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomyces.
 RX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=20036821; PubMed=10567589;
 RA Tanaka K., Nishide J., Okazaki K., Kato H., Niva O., Nakagawa T.,
 Matsuda H., Kawamukai M., Murakami Y.,
 RT "Characterization of a fission yeast SUMO-1 homologue, pmt3p,
 RT required for multiple nuclear events, including the control of
 RT telomere length and chromosome segregation.";
 RL Mol. Cell. Biol. 19:8660-8672(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=972;
 RA Wood V., Rajandream M.A., Barrell B.G., Lelaure V., Gallbert F.,
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 25-117 FROM N.A.
 RC STRAIN=358;
 RA Pelletier M.F., Dignard D.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR CHROMOSOME SEGREGATION WHERE IT MAY BE
 CC INVOLVED IN MICROTUBULE ASSEMBLY. LOSS OF SMT3 LEADS TO AN
 CC INCREASE IN TELOMERE LENGTH.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB017187; BAA32595.1; -;
 DR EMBL: AF019235; BAB71541.1; -;
 DR EMBL: AL078627; CAB44758.1; ALT_FRAME.
 DR InterPro: IPR000626; -;
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 KW Nuclear protein.
 FT DOMAIN 35 115 UBIQUITIN-LIKE.
 FT SEQUENCE 117 AA; 12934 MW; 11C860EBEA172FD2 CRC64;
 SQ

Query Match 46.6%; Score 244.5; DB 1; Length 117;
 Best Local Similarity 52.1%; Pred. No. 2.5e-17;
 Matches 50; Conservative 12; Mismatches 25; Indels 9; Gaps 1;

QY 2 SDQAKSPSTEDLGKKGQGYIKLVIGDSSSEIFKRYKMTTHLKLKESYCGQGVPMN 61
 DB 25 SQQGVKST-----EHINKVYGQDNNEVFYKIKTKTEFSKLMKICYCARQKSMNS 75

QY 62 LRFLEGGRIADNHTPKELGMEEDVEIYEOBORG 97
 DB 76 LRFLEGGRIADNHTPKELGMEEDVEIYEOBORG 111

RESULT 4
 SM31_HUMAN
 ID SM31_HUMAN STANDARD; PRT; 103 AA.
 AC P55854;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE UBIQUITIN-LIKE PROTEIN SMT3A.
 GN SMT3H1 OR SMT3A.

OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 RX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97237059; PubMed=9119407;
 RA Lapenta V., Chitrazzi P., van der Spek P.J., Pizzutti A.,
 RT Hanaoka F., Brahe C.;
 RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
 RT chromosome 21qter and defines a novel gene family.";
 RL Genomics 40:362-367(1997).
 RN [2]
 RP -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X99584; CAA67896.1; -;
 DR MIM: 602231; -;
 DR InterPro: IPR000626; -;
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 FT DOMAIN 15 92 UBIQUITIN-LIKE.
 FT SEQUENCE 103 AA; 11654 MW; E335318AC066187 CRC64;
 SQ

Query Match 45.6%; Score 239.5; DB 1; Length 103;
 Best Local Similarity 47.4%; Pred. No. 6.6e-17;
 Matches 46; Conservative 20; Mismatches 26; Indels 5; Gaps 1;

QY 1 MSDQAKSPSTEDLGKKGQGYIKLVIGDSSSEIFKRYKMTTHLKLKESYCGQGVPMN 60
 DB 1 MSEERKE-----GVKTENDHINKLVAGDGSVVOFKIRKRTSLKMKAVCERQGLSMR 55

QY 61 SLRFLFEGRIADNHTPKELGMEEDVEIYEOBORG 97
 DB 56 QIRFRFGQPINETPDAQLRMEDEDTIDVFQOQTGG 92

RESULT 5
 SM32_HUMAN
 ID SM32_HUMAN STANDARD; PRT; 95 AA.
 AC P55855;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE UBIQUITIN-LIKE PROTEIN SMT3B (SENTRIN 2).
 GN SMT3H2 OR SMT3B.
 OS Homo sapiens (Human), and Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 RX NCBI_TaxID=9606, 9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Human; TISSUE=Brain;
 RX MEDLINE=97237059; PubMed=9119407;
 RA Lapenta V., Chitrazzi P., van der Spek P.J., Pizzutti A.,
 RT Hanaoka F., Brahe C.;
 RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
 RT chromosome 21qter and defines a novel gene family.";
 RL Genomics 40:362-367(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Bovine; TISSUE=Endometrium;
 RA O'F., Berry E.S.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X9585; CAA67897.1; -
DR EMBL: U89439; AAB49682.1; -
DR MIM: 603042; -
DR InterPro: IPR000626; -
DR PROSITE: PS50053; UBIQUITIN_2; 1.
FT DOMAIN 16 95 UBIQUITIN-LIKE.
SQ SEQUENCE 95 AA: 10871 MW: 880426849BEF08B CRC64;

Query Match 44.2%; Score 232; DB 1; Length 95;
Best Local Similarity 45.4%; Pred. No. 3.2e-16;
Matches 44; Conservative 21; Mismatches 28; Indels 4; Gaps 1;

QY 1 MSDFKPSSTEDLDGKKGGEYIKLVIGODSSSEHPKVMTHLKLKESYCORGVNPN 60
DB 1 MADEPKRGV-----TENNDHNLKVAQDGSVQFKIRHPTLSKLMKAVCERQGLSMR 56
QY 61 SLRLEFGRIADNHTPKELGMEEDVIEVYQETGG 97
DB 57 QIRFRFDGPINETDTPALQLEMEDDTIDVFCQQTGG 93

RESULT 6
SMT3_ARATH STANDARD: PRT; 104 AA.
AC P55852;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE UBIQUITIN-LIKE PROTEIN SMT3.
GN SMT3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 11;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97237059; PubMed=9119407;
RA Lapenta V., Chaturazzi P., van der Spek P.J., Pizzuti A.,
RA Hanoka F., Brabe C.;
RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT chromosome 21qter and defines a novel gene family.";
RL Genomics 40:362-367(1997).

CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X95609; CAA67923.1; -
DR InterPro: IPR000626; -
DR PROSITE: PS50053; UBIQUITIN_2; 1.
FT DOMAIN 16 93 UBIQUITIN-LIKE.
SQ SEQUENCE 104 AA: 11654 MW: CDA68401C1E8028D CRC64;

Query Match 42.6%; Score 223.5; DB 1; Length 104;
Best Local Similarity 51.2%; Pred. No. 2.4e-15;
Matches 44; Conservative 14; Mismatches 25; Indels 3; Gaps 1;

QY 15 DKKGEE---YIKLVIGODSSSEHPKVMTHLKLKESYCORGVNPNSLRLEFGRI 71
DB 8 DKPGGGAHNLKVAQDGSVQFKIRHPTLSKLMKAVCERQGLSMR 67
QY 72 ADNHTPKELGMEEDVIEVYQETGG 97
DB 68 RAQRPDELDMEDGDEIDAMLHQTGG 93

RESULT 7
SMT3_ORYZA STANDARD: PRT; 100 AA.
ID SMT3_ORYZA
AC P55857;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE UBIQUITIN-LIKE PROTEIN SMT3.
GN SMT3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97237059; PubMed=9119407;
RA Lapenta V., Chaturazzi P., van der Spek P.J., Pizzuti A.,
RA Hanoka F., Brabe C.;
RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT chromosome 21qter and defines a novel gene family.";
RL Genomics 40:362-367(1997).
CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

Query Match 42.3%; Score 222; DB 1; Length 100;
Best Local Similarity 45.3%; Pred. No. 3.2e-15;
Matches 43; Conservative 19; Mismatches 29; Indels 4; Gaps 1;

QY 3 DQKAPSTEDLDGKKGGEYIKLVIGODSSSEHPKVMTHLKLKESYCORGVNPN 62
DB 6 EDKRPKPA---GGEGGAHNLKVAQDGSVQFKIRHPTLSKLMKAVCERQGLSMR 61
QY 63 RLLEFGRIADNHTPKELGMEEDVIEVYQETGG 97
DB 62 AFLFDGRRLRQGRPDLEMDGDEIDAMLHQTGG 96

RESULT 8
SMT3_YEAST STANDARD: PRT; 101 AA.
ID SMT3_YEAST
AC O12306;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE UBIQUITIN-LIKE PROTEIN SMT3.
GN SMT3 OR YDR510W OR D9719.15.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

RX NCBI_TaxID=4932;
 RA [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YPH1/YNN214;
 RA Meluh P.B., Koshland D.E.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Dietrich F.S.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NOT KNOWN. SUPPRESSOR OF MIF2 MUTATIONS.
 CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 027233; AAB01675.1;
 DR EMBL: 033057; AAB64951.1;
 DR SGD: S0002918; SMT3.
 DR InterPro: IPR000626;
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 FT DOMAIN 22 101 UBIQUITIN-LIKE.
 FT SEQUENCE 101 AA; 11597 MW; A2790DD7F315E1A7 CRC64;
 SQ
 Query Match 41.6%; Score 218.5; DB 1; Length 101;
 Best local Similarity 48.5%; Pred. No. 7.2e-15;
 Matches 49; Conservative 19; Mismatches 26; Indels 7; Gaps 3;
 QY 1 MSD-----QKRPSTEDLDGKKGKGYIKLKYIGDSSSEIFKVMTHLKKLKESYCOROG 56
 Db 1 MSSEVQKQKPKPKP--EVKPEITHMLKV-SGSSSEIFKIKITTPRLMLAFKAKRG 57
 QY 57 VPMSLRFLEGGRIADNHTPKELGMEEDVIEYQOTG 97
 Db 58 KEMDSLRFYDGRITQADOTPEDMDMDNDLIEHREQIG 98
 Db
 RESULT 9
 ID UBIQ_NPVAC STANDARD; PRT; 77 AA.
 AC P16709;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE UBIQUITIN-LIKE PROTEIN.
 GN V-UBI.
 OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OX NCBI_TaxID=46015;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90115886; PubMed=2153300;
 RA Guatino L.A.;
 RL "Identification of a viral gene encoding a ubiquitin-like protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:409-413(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=E2;
 RC MEDLINE=86115404; PubMed=3944847;
 RA Guatino L.A., Summers M.D.;
 RL "Functional mapping of a trans-activating gene required for
 RL expression of a baculovirus delayed-early gene.";
 RL J. Virol. 57:563-571(1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C6;

RX MEDLINE=94303173; PubMed=8030224;
 RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
 RL "The complete DNA sequence of Autographa californica nuclear
 RL polyhedrosis virus.";
 RC Virology 202:586-605(1994).
 CC -1- FUNCTION: UBIQUITIN MAY PLAY A ROLE IN VIRAL LIFE CYCLES, OR
 CC IN VIRUS-HOST INTERACTIONS. IT IS COVALENTLY LINKED TO COAT
 CC PROTEIN SUBUNITS OF SEVERAL DIFFERENT PLANT AND ANIMAL VIRUSES.
 CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M30305; AAA46751.1;
 DR EMBL: M37122; AAA46685.1;
 DR EMBL: L22858; AAA6665.1;
 DR PIR: A34813; UQNVAC.
 DR HSSP: P02248; IAA.
 DR InterPro: IPR000626;
 DR Pfam: PF00240; ubiquitin; 1.
 DR PRINTS: PR00348; UBIQUITIN.
 DR PROSITE: PS00299; UBIQUITIN_1; 1.
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 KW Nuclear protein; Late protein.
 FT SITE 48 48 NECESSARY FOR BRANCHED-CHAIN
 FT BINDING 76 76 MULTISUBUNITIN ADDUCTS.
 FT CONFLICT 20 20 CONFIGURATION TO ACCEPTOR PROTEINS.
 FT SEQUENCE 77 AA; 8653 MW; EF9C11D902ABC35A CRC64;
 SQ
 Query Match 15.4%; Score 81; DB 1; Length 77;
 Best local Similarity 25.0%; Pred. No. 0.13;
 Matches 13; Conservative 17; Mismatches 22; Indels 0; Gaps 0;
 QY 47 LKESYQROGVPNMSLRFLEGGRIADNHTPKELGMEEDVIEYQOTG 98
 Db 26 LKQIKADKEGVPYDQRLIFAGKQLEDSKTMADYNIQKESTLHMYRLRGY 77
 Db
 RESULT 10
 ID UBIQ_DICDI STANDARD; PRT; 76 AA.
 AC P08618;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-NOV-1991 (Rel. 20, Last annotation update)
 DE UBIQUITIN.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88152253; PubMed=2831095;
 RA Mueller-Taubenberger A., Westphal M., Jaeger E., Noegel A.,
 RA Gerisch G.;
 RL "Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-
 RL terminal tail and identification of the protein using an anti-peptide
 RL antibody.";
 RL FEBS Lett. 229:273-278(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89352609; PubMed=2548604;
 RA Ohnachi T., Giordano R., Shaw D.R., Ennis H.L.;
 RL "Molecular organization of developmentally regulated Dictyostelium
 RL discoideum ubiquitin cDNAs.";
 RL Biochemistry 28:5226-5231(1989).
 RN [3]

RP SEQUENCE FROM N.A.
 RX MEDLINE-87257921; PubMed=3037345;
 RA Giorda R., Ennis H.L.;
 RT "Structure of two developmentally regulated Dictyostelium discoideum
 RL ubiquitin genes."; Biol. 7:2097-2103(1987).
 RN [4]
 RP SEQUENCE OF 13-76 FROM N.A.
 RA STRAIN-AX2;
 RT Westphal M., Mueller-Taubenberg A., Noegel A., Gerlisch G.;
 RT "Transcript regulation and carboxyterminal extension of ubiquitin in
 RL dictyostelium discoideum."; FEBS Lett. 209:32-36(1986).
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
 CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
 CC BIOGENESIS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
 CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN
 CC PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN GENES
 CC CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X07210; CAA30183.1; ALT_TERM.
 DR EMBL: M19666; AAA33261.1; ALT_TERM.
 DR EMBL: M19491; AAA33269.1; ALT_TERM.
 DR EMBL: M19492; AAA33270.1; ALT_TERM.
 DR EMBL: M23748; AAA33262.1; ALT_TERM.
 DR EMBL: M23749; AAA33263.1; ALT_TERM.
 DR EMBL: M23750; AAA33264.1; ALT_TERM.
 DR EMBL: M23751; AAA33265.1; ALT_TERM.
 DR EMBL: M23752; AAA33266.1; ALT_TERM.
 DR EMBL: M23753; AAA33267.1; ALT_TERM.
 DR EMBL: M23754; AAA33268.1; ALT_TERM.
 DR EMBL: X04702; CAA28408.1; ALT_TERM.
 DR PIR: S00357; UOQOR.
 DR PIR: E34080; UOQOR7.
 DR PIR: A27806; A27806.
 DR PIR: B27806; B27806.
 DR PIR: A34080; A34080.
 DR PIR: B34080; B34080.
 DR PIR: C34080; C34080.
 DR HSSP: P02248; 1AAR.
 DR DictyDB: D005001; ubqa.
 DR InterPro: IPR000626;
 DR Pfam: PF00240; ubiquitin; 1.
 DR PRINTS: PR00348; UBIQUITIN.
 DR PROSITE: PS00299; UBIQUITIN_1; 1.
 DR PROSITE: PS00553; UBIQUITIN_2; 1.
 KW Nuclear protein; Polyprotein.
 FT SITE 48
 FT BINDING 76
 FT VARIANT 11
 FT SEQUENCE 76 AA; 8538 MW; 6427383968EAB84 CRC64;
 SO SEQUENCE

Query Match 15.2%; Score 80; DB 1; Length 76;
 Best Local Similarity 21.5%; Pred. No. 0.16;
 Matches 17; Conservative 20; Mismatches 30; Indels 12; Gaps 1;

QY 19 GEYIKLVYIGDSSEIHKVMTHTLKKKESYCGRGVPMNSLRFLFEQGRADNHTPK 78
 Db 10 KGTITLEVEGSD-----NIENVAKIODEKGIIPDQORLIFAGKOLEDGRTLS 57

QY 79 ELGMEEDVIEVYQETGG 97
 Db 58 DYNQKESTLHLVLRNG 76

RESULT 11
 ID UBIQ_ACACA STANDARD; PRT; 76 AA.
 AC P49634;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE UBIQUITIN.
 OS Acanthamoeba castellanii (Amoeba).
 OC Eukaryota; Acanthamoebidae; Acanthamoeba.
 OX NCBI_Taxid=5755;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-94250685; PubMed=8193157;
 RA Ahn K.S., Henney H.R.;
 RT "An Acanthamoeba ubiquitin-fusion protein; cDNA and deduced protein
 RT sequence."; Biochim. Biophys. Acta 1218:109-111(1994).
 RL Biochim. Biophys. Acta 1218:109-111(1994).
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
 CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
 CC BIOGENESIS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED WITH RIBOSOMAL CEP52 AS
 CC ITS C-TERMINAL EXTENSION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X75628; CAA53293.1; ALT_TERM.
 DR HSSP: P02248; 1AAR.
 DR InterPro: IPR000626;
 DR Pfam: PF00240; ubiquitin; 1.
 DR PRINTS: PR00348; UBIQUITIN.
 DR PROSITE: PS00299; UBIQUITIN_1; 1.
 DR PROSITE: PS00553; UBIQUITIN_2; 1.
 KW Nuclear protein; Polyprotein.
 FT SITE 48
 FT BINDING 76
 FT VARIANT 11
 FT SEQUENCE 76 AA; 8596 MW; D84480E07D1EB52 CRC64;
 SO SEQUENCE

Query Match 14.5%; Score 76; DB 1; Length 76;
 Best Local Similarity 20.3%; Pred. No. 0.39;
 Matches 13; Conservative 21; Mismatches 30; Indels 0; Gaps 0;

QY 34 IHFKVMTHTLKKKESYCGRGVPMNSLRFLFEQGRADNHTPKLGMEEDVIEVYQ 93
 Db 13 TLEVESDPIENVKQIKQDEKGIIPDQORLIFAGKOLEDGRTLDYNTQKESTLHLVLR 72

QY 94 QYQG 97
 Db 73 LRNG 76

RESULT 12
 ID NED8_HUMAN STANDARD; PRT; 81 AA.
 AC Q15843;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE UBIOUITIN-LIKE PROTEIN NEDD8.
 GN NEDD8.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fibrosarcoma;
 RA Kato S.;
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX MEDLINE=99074275; PubMed=9857030;
 RA Whitty F.G., Xia G., Pickett C.M., Hill C.P.;
 RT "Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes."
 RL J. Biol. Chem. 273:34983-34991(1998).
 CC -1- FUNCTION: ACTIVATED BY AN E1-LIKE COMPLEX, CONSISTING OF APP-B1 AND UBA3 AND THEN LINKED TO THE E2-LIKE ENZYME, UBC12. THE MAJOR TARGET PROTEIN MODIFIED BY NEDD8 IS CULLIN-4A.
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE DURING THE EMBRYONIC DEVELOPMENT AND DIFFERENTIATION OF THE CENTRAL NERVOUS SYSTEM. MAY PLAY AN ESSENTIAL ROLE IN EUKARYOTIC CELLULAR METABOLISM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE UBIOUITIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D23662; BAA04889.1; -
 CC PDB: 1NDJ; 23-FEB-99.
 DR MIM: 603171; -
 DR InterPro: IPR000626; -
 DR Pfam: PF00240; ubiqlitin; 1.
 DR PRINTS: PR00348; UBIOUITIN.
 DR PROSITE: PS00299; UBIOUITIN_1; 1.
 DR PROSITE: PS50053; UBIOUITIN_2; 1.
 DR 3D-structure.
 KW SEQUENCE 81 AA; 9072 MW; DC2FE102BE4725D2 CRC64;
 SQ
 Query Match 14.5%; Score 76; DB 1; Length 81;
 Best Local Similarity 22.4%; Pred. No. 0.42;
 Matches 17; Conservative 21; Mismatches 36; Indels 2; Gaps 1;
 QY 22 IKLVIGDSSEIHFVKMTTHLKKIKESYCORGVPMNSLRFLEGORADNHTPRELG 81
 Db 3 IKVKTLL-TGKEIIDEIPTDKVERIKERVEKEGIPPOOORLIYSKOMDEKTAADYK 60
 QY 82 MEEDVIEVYQOTGG 97
 Db 61 ILGGSVLHLVLAIRGG 76
 DB ILGGSVLHLVLAIRGG 76
 RESULT 13
 NEDD8_MOUSE STANDARD: PRT: 81 AA.
 AC P29595;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE UBIOUITIN-LIKE PROTEIN NEDD8.
 GN NEDD8 OR NEDD-8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 OX

RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93328780; PubMed=1378265;
 RA Kumar S., Tomooka Y., Noda M.;
 RT "Identification of a set of genes with developmentally down-regulated expression in the mouse brain."
 RL Biochem. Biophys. Res. Commun. 185:1155-1161(1992).
 RN (2)
 RP REVISIONS, SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93371434; PubMed=8395831;
 RA Kumar S., Yoshida Y., Noda M.;
 RT "Cloning of a cDNA which encodes a novel ubiquitin-like protein."
 RL Biochem. Biophys. Res. Commun. 195:393-399(1993).
 CC -1- FUNCTION: ACTIVATED BY AN E1-LIKE COMPLEX, CONSISTING OF APP-B1 AND UBA3 AND THEN LINKED TO THE E2-LIKE ENZYME, UBC12. THE MAJOR TARGET PROTEIN MODIFIED BY NEDD8 IS CULLIN-4A.
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE DURING THE EMBRYONIC DEVELOPMENT AND DIFFERENTIATION OF THE CENTRAL NERVOUS SYSTEM. MAY PLAY AN ESSENTIAL ROLE IN EUKARYOTIC CELLULAR METABOLISM.
 CC -1- TISSUE SPECIFICITY: UBIOUITOUS.
 CC -1- DEVELOPMENTAL STAGE: DOWN-REGULATED DURING THE DEVELOPMENT OF BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE UBIOUITIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D10918; BAA01719.1; -
 CC HSP: P02248; IAAK.
 DR MGI: 97301; Nedd8.
 DR InterPro: IPR000626; -
 DR Pfam: PF00240; ubiqlitin; 1.
 DR PRINTS: PR00348; UBIOUITIN.
 DR PROSITE: PS00299; UBIOUITIN_1; 1.
 DR PROSITE: PS50053; UBIOUITIN_2; 1.
 DR 3D-structure.
 KW SEQUENCE 81 AA; 8972 MW; DC339102BE4725D2 CRC64;
 SQ
 Query Match 14.5%; Score 76; DB 1; Length 81;
 Best Local Similarity 22.4%; Pred. No. 0.42;
 Matches 17; Conservative 21; Mismatches 36; Indels 2; Gaps 1;
 QY 22 IKLVIGDSSEIHFVKMTTHLKKIKESYCORGVPMNSLRFLEGORADNHTPRELG 81
 Db 3 IKVKTLL-TGKEIIDEIPTDKVERIKERVEKEGIPPOOORLIYSKOMDEKTAADYK 60
 QY 82 MEEDVIEVYQOTGG 97
 Db 61 ILGGSVLHLVLAIRGG 76
 DB ILGGSVLHLVLAIRGG 76
 RESULT 14
 UBIO_NEUCR STANDARD: PRT: 76 AA.
 AC P13117;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE UBIOUITIN.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-ST. LAWRENCE 74 / SL 74 / ORS 6A;
 OX

```

RX MEDLINE=89366647; PubMed=2549509;
RA Taccioli G.E., Grotewold E., Aisenberg G.O., Judewicz N.D.;
RT "Ubiquitin expression in Neurospora crassa: cloning and sequencing of
RL a polyubiquitin gene.";
RN Nucleic Acids Res. 17:6153-6165(1989).
RP
RM
RN
RP SEQUENCE FROM N.A.
RC STRAIN-ST. LAWRENCE 74 / SL 74 / ORS 6A;
RX MEDLINE=91323720; PubMed=1650731;
RA Taccioli G.E., Grotewold E., Aisenberg G.O., Judewicz N.D.;
RT "The cDNA sequence and expression of an ubiquitin-tail gene fusion in
RL Neurospora crassa.";
RN Gene 102:133-137(1991).
RP
RM
RN
RP SEQUENCE FROM N.A.
RC STRAIN-74-OR23-1A;
RX MEDLINE=94374698; PubMed=8086539;
RA Tarawneh K.A., Ammula K.R., Free S.J.;
RT "The isolation and characterization of a Neurospora crassa gene
RL (ubi::crp-6) encoding a ubiquitin-40S ribosomal protein fusion
protein.";
RN Gene 147:137-140(1994).
RP
RM
RN
RP -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN
CC PRECURSOR WITH 4 EXACT HEAD TO TAIL REPEATS. THERE IS A FINAL
CC AMINO-ACID (GLN) AFTER THE LAST REPEAT. SOME UBIQUITIN GENES
CC CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13140; CAA31530.1; ALT_TERM.
DR EMBL: U01220; AAA56880.1; ALT_TERM.
DR EMBL: U01221; AAA03351.1; ALT_TERM.
DR PIR: S05223; UNOC.
DR HSSP: P02248; IAAK.
DR InterPro: IPR000626; -
DR Pfam: PF00240; ubiquitin; 1.
DR PRINTS: PR00348; UBIQUITIN.
DR PROSITE: PS00299; UBIQUITIN_1; 1.
DR PROSITE: PS0053; UBIQUITIN_2; 1.
DR KMW Nuclear protein; Polyprotein.
FT SITE 48
FT FT
FT BINDING 76
FT SEQUENCE 76 AA; 8598 MW; 994480FE7D38403E CRC64;
SQ

```

Query Match 14.3%; Score 75; DB 1; Length 76;
Best Local Similarity 20.3%; Pred. No. 0.49;
Matches 13; Conservative 20; Mismatches 31; Indels 0; Gaps 0;

```

QY 34 IFKVKMTTHLKLKESYCOQGVPMNSLRFLFEGORIADNHTPEKLGMEEDVLEVOE 93
DB 13 IFLVESSPTIDNVKQIKDKGIPPDQORLIFAKQLEDGRTLSIDYNIQKSTLHVLVR 72
QY 94 QMG 97
DB 73 LRGG 76

```

RESULT 15
UBIO_STRPU
ID UBIO_STRPU STANDARD: PRT: 76 AA.

```

AC P23398;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last annotation update)
DE UBIQUITIN.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinozoa; Echinodermata; Echinozoa; Echinozoa; Echinozoa;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN
RP
RM
RP SEQUENCE FROM N.A.
RC TISSUE-Blastula;
RX MEDLINE=91250006; PubMed=1645680;
RA Nemer M., Rondinelli E., Infante D., Infante A.A.;
RT "Polyubiquitin RNA characteristics and conditional induction in sea
RL urchin embryos.";
RN Dev. Biol. 145:255-265(1991).
RP
RM
RN
RP -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN
CC PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. THE NUMBER OF REPEAT
CC DIFFER BETWEEN SPECIES. THERE ARE ESTIMATED TO BE APPROXIMATELY 10
CC AFTER THE LAST REPEAT.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61172; AAA30082.1; ALT_SEQ.
DR HSSP: P02248; IAAK.
DR InterPro: IPR000626; -
DR Pfam: PF00240; ubiquitin; 1.
DR PRINTS: PR00348; UBIQUITIN.
DR PROSITE: PS00299; UBIQUITIN_1; 1.
DR PROSITE: PS0053; UBIQUITIN_2; 1.
DR KMW Nuclear protein; Polyprotein.
FT SITE 48
FT FT
FT BINDING 76
FT SEQUENCE 76 AA; 8551 MW; C42D32397FED9E02 CRC64;
SQ

```

Query Match 14.3%; Score 75; DB 1; Length 76;
Best Local Similarity 21.5%; Pred. No. 0.49;
Matches 17; Conservative 19; Mismatches 31; Indels 12; Gaps 1;

```

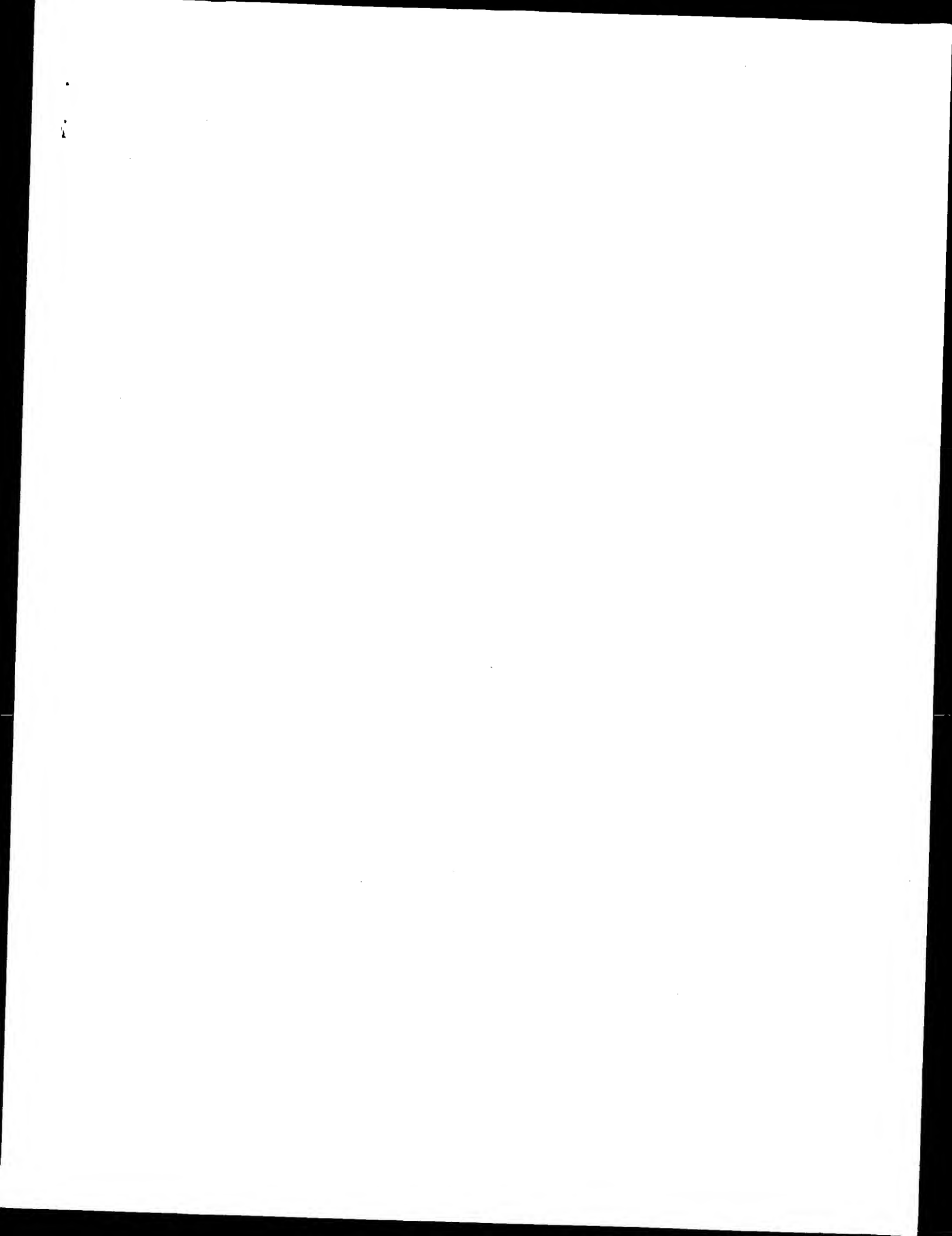
QY 19 GEYIKLVIGDSSSEIFKVKMTTHLKLKESYCOQGVPMNSLRFLFEGORIADNHTPK 78
DB 10 GKTITLEVEPDS-----IENVKAKIQDKGIPPDQORLIFAKQLEDGRTLS 57
QY 79 ELGMEEDVLEVOEQMG 97
DB 58 DYNQKSTLHVLRLRNG 76

```

Search completed: August 15, 2001, 17:18:03
Job time: 374 sec

Thu Aug 16 08:48:55 2001

us-09-484-964-2.rsp



Gencore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 17:12:59 ; Search time 91.85 seconds
(without alignments)
145.485 Million cell updates/sec

Title: US-09-484-964-2
Perfect score: 525
Sequence: 1 MSDDKAPSTEDLGDKKKE.....MEEDVIEVYQEOGTGSHSV 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_16: *
1: sp.archaea: *
2: sp.bacteria: *
3: sp.fungi: *
4: sp.human: *
5: sp.invertebrate: *
6: sp.mammal: *
7: sp.mhc: *
8: sp.organelle: *
9: sp.phage: *
10: sp.plant: *
11: sp.rodent: *
12: sp.unclassified: *
13: sp.vertebrate: *
14: sp.virus: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	517	98.5	101	6	09MZD5
2	470.5	89.6	102	13	057686
3	428	81.5	101	13	09PT08
4	239.5	45.6	110	11	09E172
5	235.5	44.9	90	5	097102
6	233	44.4	239	14	057114
7	223.5	42.6	100	10	09S224
8	221	42.1	103	10	09FLP6
9	216.5	41.2	105	10	09SMD1
10	208	39.6	115	10	023759
11	196	37.3	95	4	09S531
12	175.5	33.4	111	10	09FLP5
13	147	28.0	380	10	048841
14	137.5	26.2	114	10	09FKC5
15	120.5	23.0	117	10	09FKC6
16	104	19.8	19	11	09GVN1
17	101	19.2	412	11	009130
18	98.5	18.8	408	6	09GLZ9
19	98	18.7	156	10	09SHE7

20	97.5	18.6	459	11	09JLL1	09j111 rattus norv
21	97.5	18.6	465	11	09JMK4	09jmk4 rattus norv
22	97.5	18.6	465	11	09JMK6	09jmk6 rattus norv
23	96	18.3	78	10	063381	063381 arabidopsis
24	94.5	18.0	255	11	09ES22	09es22 mus musculu
25	94.5	18.0	262	11	09ES23	09es23 mus musculu
26	94.5	18.0	464	11	09WVS6	09wvs6 mus musculu
27	93.5	17.8	456	3	006340	006340 saccharomyc
28	92.5	17.6	465	4	060260	060260 homo sapien
29	88.5	16.9	655	5	010577	010577 caenorhabd1
30	86	16.4	84	5	09VJ33	09vj33 drosophila
31	86	16.4	154	10	09M551	09m551 populus tre
32	85.5	16.3	144	10	09SYE2	09syf2 arabidopsis
33	85.5	16.3	406	3	09USX3	09usx3 schizosach
34	83	15.8	379	5	027191	027191 tetrahymena
35	83	15.8	902	10	09FH07	09fh07 arabidopsis
36	82.5	15.7	551	10	09S118	09s118 arabidopsis
37	82	15.6	65	5	015871	015871 psalteriomo
38	82	15.6	303	5	027194	027194 tetrahymena
39	81	15.4	80	14	090195	090195 spodoptera
40	81	15.4	264	5	094813	094813 tetrahymena
41	81	15.4	379	5	027192	027192 tetrahymena
42	80	15.2	150	14	09VMT1	09vmt1 lymantiria d
43	80	15.2	349	9	09G044	09g044 bacterioph
44	80	15.2	442	1	058688	058688 pyrococcus
45	79.5	15.1	476	13	013256	013256 gallus gall

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	101 AA.
Q9MZD5	Q9MZD5			
AC	Q9MZD5			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	SENTIN.			
OS	Cervus nippon (Sika deer).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae;			
OC	Cervidae; Cervinae; Cervus.			
OX	NCBI_TaxID=9683;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Sun L.G., Yu Y.L., Jiang Y.,			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF242526; AAF97049.1; -			
DR	InterPro: IPR00626; -			
DR	Pfam: PF00240; ubiquitin.1.			
DR	PROSITE: PS50053; Ubiquitin_2; 1.			
DR	SEQUENCE 101 AA; 11566 MW; 89A1AAD2D054FB33 CRC64;			
QY	1 MSDDKAPSTEDLGDKKKEGYIKLVIGDSSSEIFPKYKMTTHLKKLESYCORGVPMN 60			
DB	1 MSDDKAPSTEDLGDKKKEGYIKLVIGDSSSEIFPKYKMTTHLKKLESYCORGVPMN 60			
QY	61 SURFLEGRIADNHTPKELGMEEDVIEVYQEOGTGSHSV 101			
DB	61 SURFLEGRIADNHTPKELGMEEDVIEVYQEOGTGSHSV 101			
RESULT	2	PRELIMINARY:	PRT:	102 AA.
ID	057686			
AC	057686;			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,
 RA Beeson K.I., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dudlin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jatali M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Kerchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Klup D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20127880; PubMed=10660560;
 RA Bhaskar V., Valentine S.A., Courey A.J.;
 RT "A functional interaction between Dorsal and components of the smt3
 RT conjugation machinery.";
 RL J. Biol. Chem. 275:4033-4040(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99060729; PubMed=9844739;
 RA Huang H.W., Tsai S.C., Sun Y.H., Li S.S.;
 RT "Identification and characterization of the smt3 cDNA and gene
 RT encoding ubiquitin-like protein from *Drosophila melanogaster*.";
 RL Biochem. Mol. Biol. Int. 46:775-785(1998).
 DR EMBL: AE003615; AAF52470.1; -;
 DR EMBL: AF218862; AAF31702.1; -;
 DR EMBL: AF053083; AAD19219.1; -;
 DR HSP: Q93068; IASR.
 DR Flybase: FBgn0026170; smt3.
 DR InterPro: IPR000626; -;
 DR PROSITE: P550053; UBIQUITIN_2; 1.
 DR SMART: SM00213; UBO; 1.
 SQ SEQUENCE 90 AA; 10124 MW; E49DA04C2A13CA5 CRC64;

Query Match 44.9%; Score 235.5; DB 5; Length 90;
 Best Local Similarity 52.3%; Pred. No. 2.2e-15;
 Matches 46; Conservative 16; Mismatches 23; Indels 3; Gaps 1;

OY 13 LGDRKGG--EYIKLVIGDSSSEIFKVKMTLKKLKSQYOCOGVPPNLSRFLFEQG 69
 DB 1 MSDEKGGEGEHNINLVAGODNAVVOFKIKRHTPLKRLIMAYCDRAGLSNQQVFRFRDGG 60
 OY 70 RIADNHTPKELGMEEDVIEVYQOTGG 97
 DB 61 PINENDPTSLMEEGDTLEVYQOQTGG 88
 RESULT 6
 ID 057114 PRELIMINARY; PRT; 239 AA.
 AC 057114;

DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE NONSTRUCTURAL PROTEIN P125-2 (FRAGMENT).
 GN P125.
 OS Muscosal disease virus.
 OC Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 NCBI_TaxID=11099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SNL;
 RA Qi F., Berry E.S.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U89438; AB92355.1; -;
 DR HSP: Q93068; IASR.
 DR InterPro: IPR000626; -;
 DR PROSITE: P550053; UBIQUITIN_2; 1.
 DR SMART: SM00213; UBO; 1.
 KW Nonstructural protein.
 FT NON_TER 1 239
 FT NON_TER 1 239
 SQ SEQUENCE 239 AA; 26881 MW; 7D7952874D37AE90 CRC64;

Query Match 44.4%; Score 233; DB 14; Length 239;
 Best Local Similarity 44.6%; Pred. No. 1.1e-14;
 Matches 45; Conservative 21; Mismatches 31; Indels 4; Gaps 1;

OY 1 MSDOEAKPSTEDIGDRKGGYIKLVIGDSSSEIFKVKMTLKKLKSQYOCOGVPPN 60
 DB 57 MADEKKEGVK---TENNDHINIKVAGDGSVQFQIKRHTPLSLKMKAYCERGLISM 112
 OY 61 SLRFLREGORINDNHTPKELGMEEDVIEVYQOTGGHSY 101
 DB 113 QIRFRDGPINETDTPPAOLEMEDDTIDVQOQTGGP 153

RESULT 7
 ID 095224 PRELIMINARY; PRT; 100 AA.
 AC 095224;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
 DE UBIQUITIN-LIKE PROTEIN.
 GN F10W23.180 OR AT4G26840.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Lechamy A., Chefor F., Krivitzky M., Kreis M.,
 RA Hohnsels J., Mewes H.W., Mayer K.F.X., Schueller C.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lechamy A., Chefor F., Krivitzky M., Kreis M., Mewes H.W.,
 RA Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL035440; CAB36530.1; -;
 DR EMBL: AL161565; CAB79539.1; -;
 DR HSP: Q93068; IASR.

DR InterPro: IPR000626; -
 DR InterPro: IPR001596; -
 DR PROSITE: PS00387; PBASE: UNKNOWN_1.
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 SQ SEQUENCE 100 AA; 10976 MW; 6F1E8721E773444A CRC64;

Query Match 42.6%; Score 223.5; DB 10; Length 100;
 Best Local Similarity 51.2%; Pred. No. 3.3e-14;
 Matches 44; Conservative 14; Mismatches 25; Indels 3; Gaps 1;

QY 15 DKRGE---YIKLKVIGDSSSEIHFVKMTTHLKKLESYCCORGVPMNSLRPLFGGORT 71
 DB 8 DKRPDGGHINLKKYGGDGNVEFFRIKSTOLKILMAYCDROSVDMSIAFLFDGRRL 67

QY 72 ADNTPKELGMEEDVIEVYQDTGG 97
 DB 68 RAQTPDELMEDGDEIDAMLHOTGG 93

RESULT 8
 Q9FLP6 PRELIMINARY; PRT; 103 AA.

AC Q9FLP6; PRT; 103 AA.
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE UBIQUITIN-LIKE PROTEIN SM73-LIKE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=96290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:41-54 (1998).
 DR EMBL; AB010071; BAB08585.1; -
 SQ SEQUENCE 103 AA; 11654 MW; E8F9C486588FF342 CRC64;

Query Match 42.1%; Score 221; DB 10; Length 103;
 Best Local Similarity 51.8%; Pred. No. 5.9e-14;
 Matches 44; Conservative 13; Mismatches 26; Indels 2; Gaps 1;

QY 15 DKR--QGEYIKLVIGDSSSEIHFVKMTTHLKKLESYCCORGVPMNSLRPLFGGORT 72
 DB 8 DKRPDGGHINLKKYGGDGNVEFFRIKSTOLKILMAYCDROSVDMSIAFLFDGRRL 67

QY 73 DNHTPKELGMEEDVIEVYQDTGG 97
 DB 68 AEQTPDELMEDGDEIDAMLHOTGG 92

RESULT 9
 Q9SMD1 PRELIMINARY; PRT; 105 AA.

AC Q9SMD1; PRT; 105 AA.
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE SUMO PROTEIN.
 GN SUMO.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
 OC Solanales; Solanaceae; Solanum.

OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hanania U., Furman-Matraso N., Ron M., Avni A.;
 RT "Isolation of a novel tomato T-SUMO protein that suppresses
 RT hypersensitive cell death.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ012717; CAB60728.1; -
 DR HSSP; Q93068; IASR.
 DR InterPro: IPR000626; -
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 DR SMART; SM00213; UBO; 1.
 SQ SEQUENCE 105 AA; 11875 MW; C8D0EA4832727779 CRC64;

Query Match 41.2%; Score 216.5; DB 10; Length 105;
 Best Local Similarity 45.3%; Pred. No. 1.6e-13;
 Matches 43; Conservative 17; Mismatches 28; Indels 7; Gaps 1;

QY 3 DQAKPSTEDLDKKRGEYIKLVIGDSSSEIHFVKMTTHLKKLESYCCORGVPMNSL 62
 DB 9 DEDKKFN-----DQVHINLKKYGGDGNVEFFRIKSTQMRKILMAYCDROSVDMSI 61

QY 63 RFLFGSGRIADNHTPKELGMEEDVIEVYQDTGG 97
 DB 62 AFLFDGRRLRAEQTPDELMEDGDEIDAMLHOTGG 96

RESULT 10
 Q23759 PRELIMINARY; PRT; 115 AA.

AC Q23759; PRT; 115 AA.
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE UBIQUITIN PROTEIN.
 OS Cicer arietinum (Chickpea) (Garbanzo).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Cicer.
 OX NCBI_TaxID=3827;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CASTELLANA; TISSUE=ETIOLATED EPICOTYLS;
 RA Munoz F.J., Dopico B., Labrador E.;
 RL Plant Physiol. 116:1605-1605(1998).
 DR EMBL; AJ001901; CAA05079.1; -
 DR HSSP; Q93068; IASR.
 DR Mendel; 26845; Cicar; 3065; 26845.

DR InterPro: IPR000626; -
 DR PROSITE: PS00387; PBASE: UNKNOWN_1.
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 DR SMART; SM00213; UBO; 1.
 SQ SEQUENCE 115 AA; 12858 MW; DE5C35328588E142 CRC64;

Query Match 39.6%; Score 208; DB 10; Length 115;
 Best Local Similarity 44.3%; Pred. No. 1.1e-12;
 Matches 43; Conservative 15; Mismatches 37; Indels 2; Gaps 1;

QY 1 MSDQAKPSTEDLDKKRGE--YIKLVIGDSSSEIHFVKMTTHLKKLESYCCORGV 58
 DB 1 MSGATPNPTPPPEEDKKENDAAHINLKKYGGDGNVEFFRIKSTOLKILMAYCDROSVD 60

QY 59 MNSLFLFEGGRIADNHTPKELGMEEDVIEVYQDTGG 95
 DB 61 LNSTAFLFDGRRLRAEQTPDELMEDGDEIDAMLHOT 97

RESULT 11
 Q95531 PRELIMINARY; PRT; 95 AA.

AC 095531;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE DJ281H8.4 PROTEIN.
 GN DJ281H8.4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mashregh, Mohammad M.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A031133; CAZ20019.1; -;
 DR HSSP: Q93068; IASR.
 DR InterPro: IPR000626; -;
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 DR SMART: SMO0213; UBO; 1.
 SQ SEQUENCE 95 AA; 10653 MW; 8C7FC2934ED23 CRC64;

Query Match 37.3%; Score 196; DB 4; Length 95;
 Best Local Similarity 44.0%; Pred. No. 1.3e-11;
 Matches 40; Conservative 19; Mismatches 30; Indels 2; Gaps 2;

QY 7 KPSTEDGDKKQGEYIKLKYIGDSSSEIHKVKMTTHLKLKESYCOROGVPMNSLFLF 66
 DB 5 KP-TEEV-KTENNNHNLKVGQDGVQVKIKQTPRLKLMKAYCEPRLSVKQIRFRF 62
 QY 67 EGORIADNHTPKELGMEEDVIEVYQOTGG 97
 DB 63 GGPISGTDKRPAQLEMEDEDTIVFOQPTGG 93

RESULT 12
 ID 09FLP5 PRELIMINARY; PRT: 111 AA.
 AC 09FLP5;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE UBIQUITIN-LIKE PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 NC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=98290546; PubMed=9628562;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., MiyaJima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned pl and TAC clones.";
 RL DNA Res. 5:41-54(1998).
 DR EMBL: AB010071; BAB08586.1; -;
 SQ SEQUENCE 111 AA; 12580 MW; 79544A00709B67EC CRC64;

Query Match 33.4%; Score 175.5; DB 10; Length 111;
 Best Local Similarity 40.4%; Pred. No. 1.3e-09;
 Matches 38; Conservative 17; Mismatches 34; Indels 5; Gaps 1;

QY 4 QEAKPSTEDGDKKQGEYIKLKYIGDSSSEIHKVKMTTHLKLKESYCOROGVPMNSLFLF 63
 DB 5 QDDKPI-----DQGEHNVHLKAKVSQDGDVEFLKKNKSAFLKMLATVYCDRSLKIDATA 59
 QY 64 FLFEGORIADNHTPKELGMEEDVIEVYQOTGG 97
 DB 60 FIFNGARIGLETPDELDMEDGDVIDACRAMSG 93

RESULT 13
 ID 048841 PRELIMINARY; PRT: 380 AA.
 AC 048841;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE F24L7.10 PROTEIN.
 GN F24L7.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 NC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
 RA Somerville C.R., Venter J.C.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC003974; AAC04487.1; -;
 DR HSSP: Q93068; IASR.
 DR InterPro: IPR000626; -;
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 DR SMART: SMO0213; UBO; 1.
 SQ SEQUENCE 380 AA; 42844 MW; C8D11F8998517C10 CRC64;

Query Match 28.0%; Score 147; DB 10; Length 380;
 Best Local Similarity 39.5%; Pred. No. 2.5e-06;
 Matches 30; Conservative 15; Mismatches 31; Indels 0; Gaps 0;
 QY 22 IKLKYIGDSSSEIHKVKMTTHLKLKESYCOROGVPMNSLFLFEGORIADNHTPKELG 81
 DB 300 VILKYNQOGADLYKIGHAHLKLMSAYCTKRNDYSSVFYNGEILKARQTPAQLH 359
 QY 82 MEEDVIEVYQOTGG 97
 DB 360 MEEDVIEVYQOTGG 375

RESULT 14
 ID 09FKC5 PRELIMINARY; PRT: 114 AA.
 AC 09FKC5;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE SIMILARITY TO SMALL UBIQUITIN-RELATED PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 NC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=98403884; PubMed=9734815;
 RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., MiyaJima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 RT Sequence features of the regions of 1,367,185 bp covered by 19
 RT physically assigned pl and TAC clones.";
 RL DNA Res. 5:203-216(1998).
 DR EMBL: AB012242; BAB09424.1; -;
 SQ SEQUENCE 114 AA; 13409 MW; 9D5C8D9154856914 CRC64;

Query Match 26.2%; Score 137.5; DB 10; Length 114;
 Best Local Similarity 38.5%; Pred. No. 5.6e-06;

2 ATGCTGACCGAGGAGCAAAACCTTCAACTGAGGACTTAGCGATAAGA 51
 17 sGlnGluTyrTlleLysLeuLysValIleGlyGlnAspSerSerGlu 34
 52 GGAAGGAGATGATTAACCTCAAAAGTTATGGACGAGATGACAGTACA 101
 34 IeHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
 102 TACATTTCAAAAGTGAATGACAAACATCTCAAGAAACCTCAAGATCA 151
 51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
 152 TACGTCAAGACAGGAGGAGTCCCAATGAATCAGTCAAGTTCTCTTGA 201
 67 uGlyGlnArgGlnLeuAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
 202 AGGTGACGAGATGCTGATATCATCTCTCCGAAAGACTGGGAGATGAGG 251
 84 IuGluAspValIleGluValTyrGlnGluGlnThrGlyLysHisSerThr 100
 252 AAGAAAGATGATTTGAATTTATCAGACAAACGCGGGGTCACTGACG 301
 101 Val 101
 302 GTT 304

seq_name: gb_est4:AA227354

seq_documentation_block:

LOCUS AA227354 361 bp mRNA EST 24-FEB-1997
 DEFINITION z122612.r1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
 OF MIP2 MUTATIONS. ; similar to TR:G927779 G927779 SUPPRESSOR
 OF MIP2 MUTATIONS. ; mRNA sequence.

ACCESSION AA227354
 VERSION AA227354.1 GI:1848909
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 361)

AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldi, M.F., Chapelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
 M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore,
 B., Morris, M., Parsons, D., Prange, C., Riklin, L., Rohlfing, T.,
 Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, D., Trevisan, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

COMMENT

contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28m3 rev1 ET from Amerisham
 High quality sequence stop: 282.

FEATURES

source

1..361
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="GDB:5426089"
 /db_xref="taxon:9606"
 /clone="IMAGE:664172"
 /clone_lib="Stratagene NT2 neuronal precursor 937230"
 /tissue_type="neuroepithelial cells"
 /dev_stage="Ntera-2 neuroepithelial cells"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: Brain; Vector: pBluescript SK-; site_1:
 EcoRI; site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Uninduced, exponentially growing neuroepithelial

cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG
 3' -3' adaptor sequence: 5' CTCAGATTTTCTTTTCTTTT 3'

alignment_scores:
 Quality: 522.00 Length: 101
 Ratio: 5.168 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
 US-09-484-964-2 x AA227354

Align seq 1/1 to: AA227354 from: 1 to: 361

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLys 17
 54 ATGCTGACCGAGGAGCAAAACCTTCAACTGAGGACTTAGCGATAAGA 103
 17 sGlnGluTyrTlleLysLeuLysValIleGlyGlnAspSerSerGlu 34
 104 GGAAGGAGATGATTAACCTCAAAAGTTATGGACGAGATGACAGTACA 153
 34 IeHisPheLysValLysMetThrThrHisLeuLysLysLeuGluSer 50
 154 TACATTTCAAAAGTGAATGACAAACATCTCAAGAAACCTCAAGATCA 203
 51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
 204 TACGTCAAGACAGGAGGAGTCCCAATGAATCAGTCAAGTTCTCTTGA 253
 67 uGlyGlnArgGlnLeuAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
 254 GGTGACGAGATGCTGATATCATCTCTCCGAAAGACTGGGAGATGAGG 303
 84 IuGluAspValIleGluValTyrGlnGluGlnThrGlyLysHisSerThr 100
 304 AAGAAAGATGATTTGAATTTATCAGACAAACGCGGGGTCACTGACG 353
 101 Val 101
 354 GTT 356

seq_name: gb_est4:AA227355

seq_documentation_block:

LOCUS AA227355 368 bp mRNA EST 11-MAR-1998
 DEFINITION z122612.r1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
 OF MIP2 MUTATIONS. ; mRNA sequence.

ACCESSION AA227355
 VERSION AA227355.1 GI:1848910
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 368)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
 J., Moore, B., Scheinberg, K., Steptoe, M., Tan, F., Theising, B.,
 White, Y., Wylie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project

COMMENT

Unpublished (1997)
 contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 1260 Std Error: 0.00
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 360.

FEATURES

Source

1..368
 /organism="Homo sapiens"
 /db_xref="GDB:5426092"
 /db_xref="taxon:9606"
 /clone="IMAGE:664174"
 /clone_lib="Stratagene NT2 neuronal precursor 937230"
 /tissue_type="neuroepithelial cells"
 /dev_stage="Ntera-2 neuroepithelial cells"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: brain; Vector: pBluescript SK-; Site: 1;
 EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
 oligo dt. Uninduced, exponentially growing neuroepithelial
 cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
 Uni-ZAP XR Vector. -5' adaptor sequence: 5' GAATTCGCGACGAC
 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 122 a 69 c 91 g 86 t
 ORIGIN

alignment_scores: Length: 101
 Quality: 522.00 Gaps: 0
 Ratio: 5.168
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
 US-09-484-964-2 x AA227355 ..

Align seg 1/1 to: AA227355 from: 1 to: 368

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
53 ATGCTCTACCAAGGAGCAAACTTCACTGACGAGCTGGGGGATTAAGAA 102
17 sGlnGlyGluTyrTlleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
103 GGAAGGAGTAATATATTAACCTCAAGTCAATGGACAGATGAGTGAAGA 152
34 LeuHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
153 TTCACTTCAAAAGTCAAAATGACAAACATCTCAAGAAATCAAAAGATCA 202
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
203 TACTGTCAAAAGACAGGGGTTCCTCAATGAATTCATCAGGTTCTCTTGA 252
67 uGlyGlnArgTlleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
253 GGGTCAGAGAAATGCTGATTAATCACTACTCCAAAGAACTGGGAATGAG 302
84 LuGlnAspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThr 100
|||||
303 AAGAAGATGTGATTGAAGTTTATGCAAGCAAGGGGGGTCAATCAACA 352
101 Val 101
||||
353 GTT 355
```

seq_name: gb_est17:AI180895

seq_documentation_block:

LOCUS AI180895 388 bp mRNA EST 08-OCT-1998
 DEFINITION ub77a05.t1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
 IMAGE:1383728 5' similar to SW:SM33_HUMAN Q93068 UBIQUITIN-LIKE
 PROTEIN SWT3C ; mRNA sequence.

ACCESSION AI180895
 VERSION AI180895.1 GI:3731533
 KEYWORDS EST.
 SOURCE house mouse.

ORGANISM

REFERENCE

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 388)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 CONTACT: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:906196

COMMENT

Possible reversed clone: similarity on wrong strand
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 185.

FEATURES

Source

1..388
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:1383728"
 /clone_lib="Soares_mammary_gland_NMLMG"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="IDH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from mammary
 gland tissue from a lactating female, and was then primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT73 vector. Library is normalized. Library
 was constructed by Bento Soares and M. Fatima Bonaldi."

BASE COUNT 129 a 72 c 103 g 82 t 2 others
 ORIGIN

alignment_scores: Length: 101
 Quality: 522.00 Gaps: 0
 Ratio: 5.168
 Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
 US-09-484-964-2 x AI180895 ..

Align seg 1/1 to: AI180895 from: 1 to: 388

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
|||||
86 ATGCTGACCAAGGAGCAAACTTCACTGACGAGCTTGAAGCGTAAAGAA 135
17 sGlnGlyGluTyrTlleLysLeuLysValIleGlyGlnAspSerSerGlu 34
|||||
136 GGAAGGAGTAATATTAACCTCAAGTCAATGGACAGATGAGTGAAGA 185
34 LeuHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
|||||
186 TACATTCAAAAGTCAAAATGACAAACATCTCAAGAAATCAAAAGATCA 235
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
|||||
236 TACTGTCAAAAGACAGGGGTTCCTCAATGAATTCATCAGGTTCTCTTGA 285
67 uGlyGlnArgTlleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84
|||||
```


(pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT713 vector library. This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:671589

BASE COUNT 134 a 74 c 105 g 89 t
ORIGIN

alignment_scores:
Quality: 522.00 Length: 101
Ratio: 5.168 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
US-09-484-964-2 x AA421208

Align seg 1/1 to: AA421208 from: 1 to: 402

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
93 ATGCTGACACGAGGAGGCAAAACCTTCACTGAGGACTGGGGATTAAGAA 142
17 scInGlyGlyTyrIleLysLeuValIleGlyGlnAspSerSerGlu 34
  |||
143 GGAAGGTGATATATTAATCAAGATCATGGACAGATAGCAGTGAGA 192
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
  |||
193 TTCACCTTCAAGTGAATGACACACATCTCAAGAAATCAAGAAATCA 242
51 TyrCysGlnArgGlnGlyValProMetLysSerLeuArgPheLeuPheG1 67
  |||
243 TACTGTCAAGACAGAGGTGTTCCAAATTCACATCTGAGTTCTCTTGA 292
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
  |||
293 GGGTCGACGAATTTGCTGATATCACTCCAAAGAACTGGGAATGGAGG 342
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyLysHisSerThr 100
  |||
343 AAGAAATGTGATTTGAAGTTATTCAGAAACAAACGGGGGTCATTCACA 392
101 Val 101
  |||
393 GTT 395

```

seq_name: gb_est13:AA896203

seq_documentation_block: 426 bp mRNA EST 06-APR-1998
LOCUS AA896203
DEFINITION v62d07.r1 StrataGene mouse macrophage (#937306) Mus musculus cDNA
clone IMAGE:1279789 5' similar to SW:SM33_HUMAN Q93068
UBIQUITIN-LIKE PROTEIN SMT3C ; mRNA sequence.

ACCESSION AA896203
VERSION AA896203.1 GI:3032596
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 426)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:671589
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 407.
Location/Qualifiers

FEATURES
source

```

1..426
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1279789"
/clone_lib="Stratagene mouse macrophage (#937306)"
/tissue_type="macrophage"
/dev_stage="WEHI-3 cell line"
/lab_host="SOLR (kanamycin resistant)"
/note="organ: blood; Vector: pBluescript SK-; Site: 1:  
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT; WEHI-3 cell line. Average insert size: 1.5 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG  
3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTT 3'"
BASE COUNT 139 a 80 c 114 g 93 t
ORIGIN

```

alignment_scores:
Quality: 522.00 Length: 101
Ratio: 5.168 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.010

alignment_block:
US-09-484-964-2 x AA896203

Align seg 1/1 to: AA896203 from: 1 to: 426

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
116 ATGCTGACACGAGGAGGCAAAACCTTCACTGAGGACTGGGGATTAAGAA 165
17 scInGlyGlyTyrIleLysLeuValIleGlyGlnAspSerSerGlu 34
  |||
166 GGAAGGTGATATATTAATCAAGATCATGGACAGATAGCAGTGAGA 215
34 LeHisPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSer 50
  |||
216 TACATTTCAAGGAAATGACACACATCTCAAGAACTCAAGAAATCA 265
51 TyrCysGlnArgGlnGlyValProMetLysSerLeuArgPheLeuPheG1 67
  |||
266 TACTGTCAAGACAGAGGAGTTCATGAATTCATCAGAGTTCTCTTGA 315
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
  |||
316 AGGTCAGAGATTTGCTGATATCACTCCGAAAGAACTGGGAATGGAGG 365
84 LuGluAspValIleGluValTyrGlnGlnGlnThrGlyLysHisSerThr 100
  |||
366 AAGAAATGTGATTTGAAGTTATTCAGAAACAAACGGGGGTCATTCAGC 415
101 Val 101
  |||
416 GTT 418

```

seq_name: gb_est15:AI047190

seq_documentation_block: 429 bp mRNA EST 08-JUL-1998
LOCUS AI047190
DEFINITION uh62d04.r1 Soares-embryonic_stem_cell MNES Mus musculus cDNA clone
IMAGE:1749991 5' similar to SW:SM33_HUMAN Q93068 UBIQUITIN-LIKE
PROTEIN SMT3C ; mRNA sequence.
ACCESSION AI047190
VERSION AI047190.1 GI:3295477
KEYWORDS EST.

WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:488791
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 422.

FEATURES

source

1..434

Location/Qualifiers

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:820511"
/clone_lib="Soares_mammary_gland_NBMNG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland. Vector: p773D-Pac (Pharmacia)
with a modified polylinker. Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5']
TGTTCACATCTGAGTGGAGCGCGCGAATGTTTCTTTTCTTTTCTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT 138 a 82 c 102 g 112 t
ORIGIN

alignment_scores:

Quality: 522.00 Length: 101
Ratio: 5.168 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.010

alignment block:

US-09-484-964-2 x AAA423388 ..

Align seg 1/1 to: AAA423388 from: 1 to: 434

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
83 ATGTCGACGACGAGGCAAAACCTTCACCTAGGCGGATTAAGAA 132
17 scInGlyGluTyrIleLysLeuLysValIleGlyLnaSpSerSerGluI 34
  |:::|
133 GCAAGAGAGATACATTAACTCAAGTTATTGGACAGATGACACTGAGA 182
34 leHisPheLysValLysMetThrThrHisLysLysLysLysGluSer 50
  |||
183 TACATTTCAAAGTGAATGACACACATCTCAAGAACTCAAAAGAAATCA 232
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG 67
  |||
233 TACTGTCAAGACAGGAGCTTCACATGAAATCACCACAGTTCTTCTTGA 282
67 uGlyLnaArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
  |||
283 AGGTCAGAGAAATTCGTGATTAATCATATCTCCGAAAGAACTGGGAATGAGG 332
84 luGluAspValIleGluValTyrGlnGluGluThrGlyLysHisSerThr 100
  |||
333 AAGAGAGATGATGATTGAAGTTATCATGAGAACAAACGGGGGTCCTCGACG 382
101 Val 101
  |||

```

383 GTT 385

seq_name: gb_est23:AI646104

seq_documentation_block:

LOCUS AI646104 438 bp mRNA

DEFINITION U077405.X1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone

IMAGE:1383728.3 similar to SW:SM33_HUMAN Q93068 DB100ITIN-LIKE

PROTEIN SM33C ; mRNA sequence.

ACCESSION AI646104

VERSION AI646104.1 GI:4724579

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 438)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

COMMENT Unpublished (1997)

CONTACT Contact: Robert Strausberg, Ph.D.

EMAIL Email: cgapsb@mail.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:906196

This clone was previously sequenced on the 5' end only, this new

data is from the 3' end

Possible reversed clone: similarity on wrong strand

High quality sequence stop: 389.

Location/Qualifiers

1..438

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:1383728"

/clone_lib="Soares_mammary_gland_NMLMG"

/sex="female (lactating)"

/tissue_type="mammary gland"

/lab_host="DH10B"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker; 1st strand cDNA was prepared from mammary

FEATURES

source

1..438

Location/Qualifiers

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:1383728"

/clone_lib="Soares_mammary_gland_NMLMG"

/sex="female (lactating)"

/tissue_type="mammary gland"

/lab_host="DH10B"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker; 1st strand cDNA was prepared from mammary

gland tissue from a lactating female, and was then primed

with a Not I - oligo(dT) primer. Double-stranded cDNA was

ligated to Eco RI adaptors (Pharmacia), digested with Not

I and cloned into the Not I and Eco RI sites of the

modified p773 vector. Library is normalized. Library

was constructed by Bento Soares and M. Fatima

Bonaldo."

BASE COUNT 127 a 86 c 68 g 157 t

ORIGIN

alignment_scores:

Quality: 522.00 Length: 101

Ratio: 5.168 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.010

alignment block:

US-09-484-964-2 x AI646104/rev ..

Align seg 1/1 to reverse of: AI646104 from: 1 to: 438

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
438 ATGTCGACGACGAGGCAAAACCTTCACCTAGGCGGATTAAGAA 389
17 scInGlyGluTyrIleLysLeuLysValIleGlyLnaSpSerSerGluI 34
  |:::|
388 GCAAGAGAGATACATTAACTCAAGTTATTGGACAGATGACAGTGA 339
34 leHisPheLysValLysMetThrThrHisLysLysLysLysGluSer 50
  |||
338 TACATTTCAAAGTGAATGACACACATCTCAAGAACTCAAAAGAAATCA 289

```


51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67
 |||
 288 TACTGTCAAGACAGAGGAGTTCACATGATTCACCTGAGTTCTCTTGA 239
 |||
 67 uG1yGlnArg1leAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
 |||
 238 AGCTACAGCAATTCGTCTATATCATCTCCGAAGAAGCTGGGATGAGG 189
 |||
 84 luG1uAspVal11leGluVal1yTrGlnGluGlnThrGlyGlyHisSerThr 100
 |||
 188 AAGAGATGTGATTCAGATTATCAGCAACAAAGCGGGGCTCATCTGACG 139
 |||
 101 Val 101
 |||
 138 GTT 136
 |||
 seq_name: gb_est66:BF323866
 seq_documentation_block:
 LOCUS BF323866 439 bp mRNA EST 21-NOV-2000
 DEFINITION SNEST420f08.y1 csn 1 s neuraona invitro mezozoite CDNA Sarcocystis
 neuraona CDNA 5' similar to SW:SM33_HUMAN O93068 UBIQUITIN-LIKE
 PROTEIN SMT3C ; mRNA sequence.
 ACCESSION BF323866
 VERSION BF323866.1 GI:11273469
 KEYWORDS EST.
 SOURCE Sarcocystis neuraona.
 ORGANISM Sarcocystis neuraona
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
 Sarcocystidae; Sarcocystis.
 REFERENCE 1 (bases 1 to 439)
 HOWE,D.K., STAMPER,S., TANG,K., SHIPLEY,L.D., CLIFFON,S., MARRA,M.,
 HILLIER,L., PAPE,D., MARTIN,J., WYLIE,T., THEISING,B., BOWERS,Y.,
 GIBBONS,M., RITTER,E., MCCANN,R., BLISTAIN,A., BENNETT,J., SCHMITT,
 A., KONKO,I., TSAGAREISHVILI,R., FEDELE,M., BELAYGOROD,L.,
 FRANKLIN,C., CARR,L.M., GROW,A., MAGUIRE,L., WADKINS,J., RICHIEY,J.,
 WATERSTON,R. and WILSON,R.
 Sarcocystis neuraona EST Project
 Unpublished (2000)
 JOURNAL Contact: Daniel K. Howe
 COMMENT Sarcocystis neuraona EST project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Contact Daniel K. Howe (dkhowe2@pop.wky.edu) for further
 information relating to organism, libraries, or clone availability.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 376.
 Location/Qualifiers
 1..439
 /organism="Sarcocystis neuraona"
 /strain="Sn3"
 /db_xref="taxon:42890"
 /clone_lib="csn 1 s neuraona invitro mezozoite CDNA"
 /dev_stage="mezozoite"
 /lab_host="DH10B"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI. The library was constructed by Dan Howe, University
 of Kentucky. cDNAs were synthesized from poly(A)+ RNA
 by oligo d(T) priming and directionally cloned into the
 Uni-ZAP XR lambda vector. The library was mass excised
 as phagemids and rescued in SOLR cells. The plasmid
 library was recovered from the SOLR cells and transformed
 in mass into DH10B cells for sequencing. WARNING: the
 library contains a small percentage of cDNAs derived from
 the bovine host cells."
 BASE COUNT 139 a 81 c 98 g 121 t
 ORIGIN
 alignment_scores:
 51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67
 |||
 288 TACTGTCAAGACAGAGGAGTTCACATGATTCACCTGAGTTCTCTTGA 239
 |||
 67 uG1yGlnArg1leAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
 |||
 238 AGCTACAGCAATTCGTCTATATCATCTCCGAAGAAGCTGGGATGAGG 189
 |||
 84 luG1uAspVal11leGluVal1yTrGlnGluGlnThrGlyGlyHisSerThr 100
 |||
 188 AAGAGATGTGATTCAGATTATCAGCAACAAAGCGGGGCTCATCTGACG 139
 |||
 101 Val 101
 |||
 138 GTT 136
 |||

Quality: 522.00 Length: 101
 Ratio: 5.168 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.010
 alignment_block:
 US-09-484-964-2 x BF323866 ..
 Align seg 1/1 to: BF323866 from: 1 to: 439
 1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
 |||
 62 ATGCTGACCAAGAGAGCAAAACCTTCACACGAGACTGGGGGATTAAGA 111
 |||
 112 GGAGAGAGGTATATTAACTCAAGTCATGACAGAGATGACAGTGA 161
 |||
 17 sGlnGlyGluTyr11leLysLeuLysVal11leGlyGlnAspSerSerLut 34
 |||
 34 leHisPheLysValLysMetThrThrHisLeuLysLysLeuGlySer 50
 |||
 162 TTCACCTCAAACTGAAATGACACACATCTCAAGAAACCTCAAGAAATCA 211
 |||
 51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67
 |||
 212 TACTGTCAAGACAGAGGAGTTCACATGATTCACCTGAGTTCTCTTGA 261
 |||
 67 uG1yGlnArg1leAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
 |||
 262 AGCTACAGCAATTCGTCTATATCATCTCCGAAGAAGCTGGGATGAGG 311
 |||
 84 luG1uAspVal11leGluVal1yTrGlnGluGlnThrGlyGlyHisSerThr 100
 |||
 312 AAGAGATGTGATTCAGATTATCAGCAACAAAGCGGGGCTCATCTGACG 361
 |||
 101 Val 101
 |||
 362 GTT 364
 |||
 seq_name: gb_est10:AA636685
 seq_documentation_block:
 LOCUS AA636685 442 bp mRNA EST 22-OCT-1997
 DEFINITION v15f102.r1 Barsstead mouse myotubes MRLB5 Mus musculus CDNA clone
 IMAGE:1120731 5' similar to TR:G927779 G927779 SUPPRESSOR OF M12
 MUTATIONS. ; mRNA sequence.
 ACCESSION AA636685
 VERSION AA636685.1 GI:2560464
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 442)
 MARRA,M., HILLIER,L., ALLEN,M., BOWLES,M., DIETRICH,N., DUBUQUE,T.,
 GEISEL,S., KUOABA,T., LACY,M., LE,M., MARTIN,J., MORRIS,M.,
 SCHEIDENBERG,K., STEPTOE,M., TAN,F., UNDERWOOD,K., MOORE,B.,
 THEISING,B., WYLIE,T., LEMONN,G., SOARES,B., WILSON,R. and
 WATERSTON,R.
 The WashU-HHMI Mouse EST Project
 Unpublished (1996)
 JOURNAL Contact: Maria M/Mouse EST Project
 COMMENT WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAG Consortium (info@image.llnl.gov) for further information.
 MG1:610067
 Seq primer: -28m3 rev2 ET from Amersham
 High quality sequence stop: 441.
 Location/Qualifiers
 1..442
 FEATURES
 source


```

/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone_image="1120731"
/clone_lib="Barstead mouse myotubes MPRB5"
/cell_line="C2C12"
/lab_host="DH10B"
/notes="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Site 1: EcoRI; Site 2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCGATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pRTT3 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins."

```

```

BASE COUNT      142 a      82 c      99 g      119 t
ORIGIN

```

```

alignment_scores:
  Quality: 522.00
  Ratio: 5.168
  Percent Similarity: 100.000
  Percent Identity: 99.010

```

```

alignment_block:
  US-09-484-964-2 x AA636685
  Align seg 1/1 to: AA636685 from: 1 to: 442

```

```

1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
78 ATGTCTGACACGAGAGCAAAACCTTCACTGAGACTAGCGATTAAGAA 127
  |||
17 scngllyglutyrilleylsleuysvalileglynaspserserGIut 34
  |||
128 GGAAGAGAGATACATTAACCTTAAGTTATGAGACAGATGACAGTGACA 177
  |||
34 leHisPheLysValLysMetThrThrlsleuLysLysLysLysLysLys 50
  |||
178 TACATTTCAAGTGAAGTGAACACACATCTCAGAAACTCAAAAGATCA 227
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67
  |||
228 TACTGTCAAGACAGGAGGTTCCATGATCACTCAGGTTCTCTTTGA 277
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
  |||
278 AGGTCAAGAGAAATGCTGATATCATACCCGAAAGAACTGGGAATGAGG 327
  |||
84 lUGluAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThr 100
  |||
328 AAGAAGATGTGATTAAGTTATCAGGAACAAACGGGGGTCACTCAGACG 377
  |||
101 Val 101
  |||
378 GTT 380

```

```

seq_name: gb_esc75:BE482568

```

```

seq_documentation_block:
  LOCUS      BE482568      444 bp      mRNA      EST      28-AUG-2000
  DEFINITION 168404 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
  ACCESSION  BE482568
  VERSION    BE482568.1 GI:9602101
  KEYWORDS   EST.
  SOURCE     EST.
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
             Bovidae; Bos.

```

```

REFERENCE      1 (bases 1 to 444)
AUTHORS      Sonstegard,T.S., Capuco,A.V., Van Tassel,J.C.P., Ashwell,M.S. and
              Wells,R.D.
TITLE        Mapping of Expressed Sequence Tags from a normalized bovine mammary
              gland cDNA library
JOURNAL      Unpublished (2000)
COMMENT      Contact: Sonstegard TS
              USDA, ARS, Beltsville Agricultural Research Center
              Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
              Tel: 301 504 8416
              Fax: 301 504 8414
              Email: tads@psi.barc.usda.gov

```

```

Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers

```

```

FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 11 row: C column: 20
Seq primer: ATTAGGTGACACTATG.

```

```

FEATURES
  source
    Location/Qualifiers
      1..444
        /organism="Bos taurus"
        /db_xref="taxon:9913"
        /clone_lib="BARC 5BOV"
        /tissue_type="pooled"
        /lab_host="DH10B"
        /note="Vector: PCMV SPORT6; Site 1: XbaI; Site 2: XhoI;
        library made from pooled mRNA isolated from mammary
        tissues at eight physiological, developmental, and disease
        states."

```

```

BASE COUNT      139 a      86 c      107 g      112 t
ORIGIN

```

```

alignment_scores:
  Quality: 522.00
  Ratio: 5.168
  Percent Similarity: 100.000
  Percent Identity: 99.010

```

```

alignment_block:
  US-09-484-964-2 x BE482568
  Align seg 1/1 to: BE482568 from: 1 to: 444

```

```

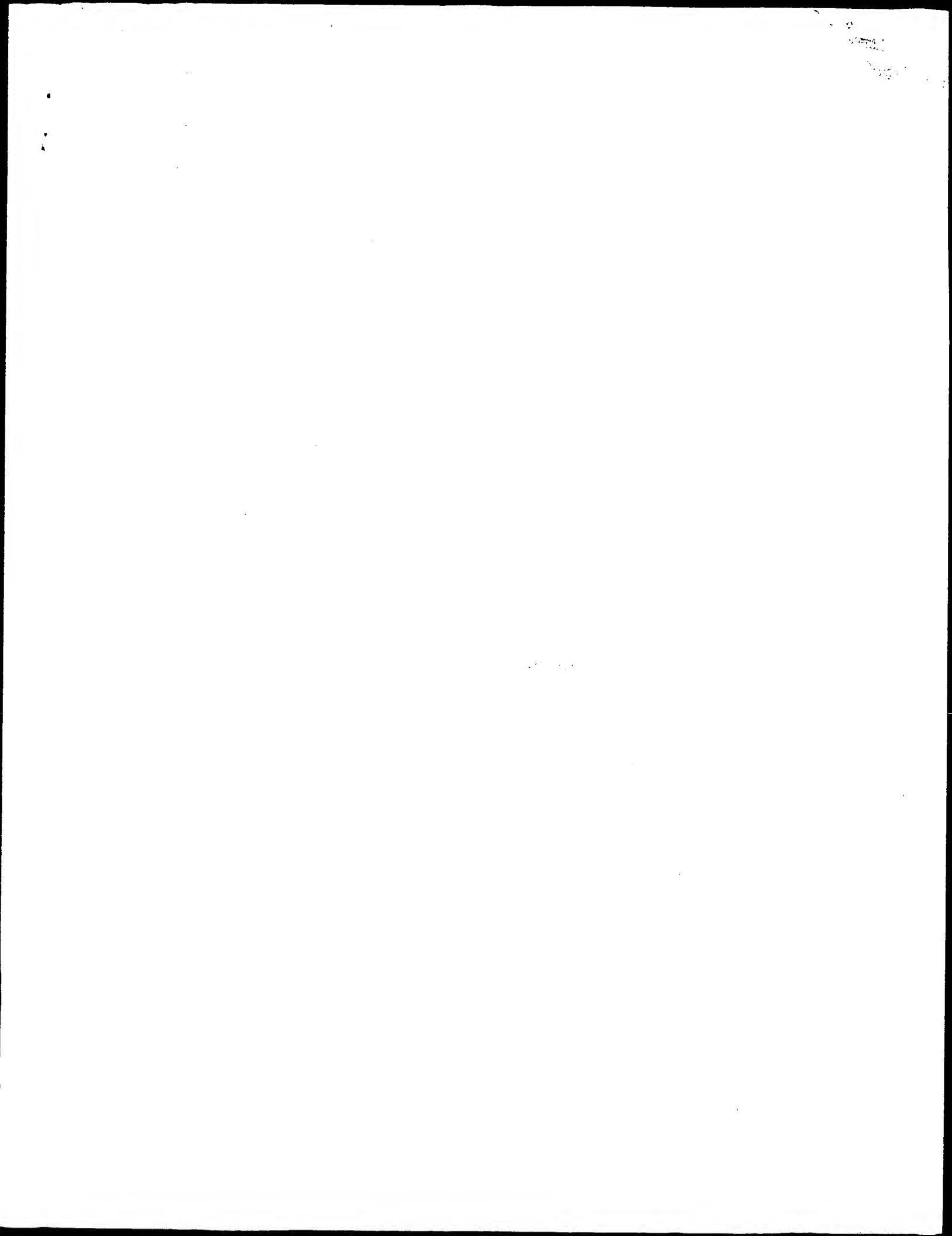
1 MetSerAspGlnGluAlaLysProSerThrGluAspLeuGlyAspLysLys 17
  |||
96 ATGTCTGACACGAGAGCAAAACCTTCAACGAGACTTGGGATTAAGAA 145
  |||
17 scngllyglutyrilleylsleuysvalileglynaspserserGIut 34
  |||
146 GGAAGAGAGATATTAAGTCAAAAGTCATTGAGACAGATGACAGTGACA 195
  |||
34 leHisPheLysValLysMetThrThrlsleuLysLysLysLysLysLys 50
  |||
196 TTCACTTCAAGTGAAGTGAACACACATCTCAAGAACTCAAAAGATCA 245
  |||
51 TyrCysGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67
  |||
246 TACTGTCAAGACAGGAGGTTCCATGATCACTCAGGTTCTCTTTGA 295
  |||
67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluG 84
  |||
296 AGGTCAAGAGAAATGCTGATATCATACCTCAGAAAGAACTGGGAATGAGG 345
  |||
84 lUGluAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThr 100
  |||
346 AAGAAGATGTGATTAAGTTATCAGGAACAAACGGGGGTCACTCAGACG 395
  |||
101 Val 101
  |||
396 GTT 398

```

Thu Aug 16 08:48:56 2001

us-09-484-964-2.rst

Page 11



OM of: US-09-484-964-2 to: GenEmbl:* out_format : pfs

Date: Aug 15, 2001 6:39 PM

About: Results were produced by the GenCore software, version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```

ncrna-11name=spn.model -DEV-xip
o/cgna2.1/OSPIO.spool/US09484964/rnaut.14082001.111933.27689/app_query.fasta.1.156
-b=GenEnb1l -qmt=fastap -SUFFIX=011.rge -GADP=4.500
-GADP=0.050 -MINMATCH=0.100 -LOOPC=0.000 -LOOPEXT=0.000
-GADP=4.500 -GADPEXT=0.050 -XGADP=60.000 -XADPEXT=60.000
-GADP=6.000 -FGADPEXT=7.000 -YGADP=60.000 -YADPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=011g
-TRANS=humana40.cdi -LIST=45 -DCCALIGN=200 -THR_SCORE=quality
-THR_MIN=20 -ALIGN=15 -MODE=LOCAL -OUTFM=score -MILEN=0
-MAXLEN=20000000 -USER=US09484964.ecgn1.14687 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLXPY -WAIT -THREADS=1

```

Search information block
Query: PE-00-404 004 0

```
Query length: 101
Database: GenEmbl:*
Database sequences: 1344157
Database length: -85606004
Search time (sec): 1073.880000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list

seq_name	gb_pr10:HSU72722	Strd	Orig	Zscore	Escore	Len	Documentation
gb_pr10:HSU72722			83.00	1639.34	4.5e-83	306	U72722 Human gap modifying pro
gb_pr9:HSMS3C		+	83.00	1634.68	8.1e-83	590	X9566 H.sapiens mRNA for SMT33
gb_pr10:HSU67122		+	83.00	1632.38	1.7e-82	816	U67122 Human ubiquitin-related
gb_pr10:HSU33874		+	83.00	1630.81	1.3e-82	1017	U38784 Human ubiquitin-like p
gb_pr10:HSU051397		+	83.00	1629.71	1.5e-82	1187	AF033353 Mus musculus ubiqit
gb_pr7:RC0084652		+	83.00	1629.50	1.6e-82	1223	U61397 Human ubiquitin-homolo
gb_pr10:HSU083117		+	83.00	1629.48	1.6e-82	1227	BC006462 Homo sapiens, ubiqutin
gb_pr10:HSU083117		+	83.00	1629.98	1.9e-82	1514	AR087051 Sequence 1 from paten
gb_om:AF242526		+	73.00	1436.90	1.9e-82	1514	U83117 Human sentrin mRNA, se
gb_pr10:HSU021844		+	67.00	1319.85	8.5e-72	510	AF242526 Cervus nippon sentrin
gb_pr10:AC084066		+	43.00	799.78	1.6e-65	372	AF021844 Sequence 18 from Paten
gb_pr10:AC153282		+	43.00	799.78	1.6e-66	235411	AC084066 Mus musculus clone
gb_pr10:AL155981		+	43.00	799.78	1.6e-66	155913	AL153282 Homo sapiens chrom
gb_pr5:AL153926		+	42.00	779.98	3.4e-36	167440	AL155981 Homo sapiens chrom
gb_pr3:AC011450		+	33.00	603.70	2.2e-25	185819	AL153926 Homo sapiens chrom
gb_pr10:AC086786		+	33.00	600.00	3.2e-25	106168	AC011450 Homo DNA sequence
gb_ov:MSU00		+	31.00	602.50	2.5e-25	178733	AC011450 Homo sapiens chrom
gb_pr10:AC079354		+	26.00	461.57	2.5e-25	466	Z97073 Xenopus laevis SMO-1 mR
gb_ov:AB036430		+	20.00	361.22	1.8e-17	160939	AC079354 Homo sapiens chrom
						670	AB036430 Oncorhynchus mykiss S

```
seq_documentation_block
```

DEFINITION	Human gap modifying protein 1 mRNA, complete cds.
ACCESSION	U72722
VERSION	U72722.1
KEYWORDS	GI:1703502
SOURCE	human.

ORGANIZATIONAL

REFERENCE
1 (bases 1 to 306)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE

TABLE 1

J. Cell Biol. 135 (6 Pt 1), 1457-1470 (1996)

JOURNAL

MEDLINE	97133418
REFERENCE	2 (bases 1 to 306)
AUTHORS	Matunis M.J., Coutavas, E. and Blobel, G.
TITLE	Direct Submission
JOURNAL	Submitted (27-SEP-1996) Laboratory of Cell Biology, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA
FEATURES	Location/Qualifiers

Source

CDS

```

CDS
1..306
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Hela"
1..306
/note="GMP1: covalently associated with Ran GTPase
activating protein RangAP1, ubiquitin-like protein"
/codon_start=1
/product="gap modifying protein 1"
/protein_id="AAB40390.1"
/db_xref="GI:17035037"
/translation="MSDQKAPSTEDGDKKEGEYIKLVIGDSSEIHFVKYMTTH
KKKSTQGRGVWNSLRFLEBSORLADNHPPKEIGMEEDVIVFYDQGTGSHSTV

```

BASE COUNT	112 a	48 c	75 g	71 t
ORIGIN				

alignment_scores:

Quality:	83.00	Length:	83
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

```
alignment_block;
```

US-09-484-964-2 X HSU72722

Align seg 1/1 to: HSU72722 from: 1 to: 306

```

13  G1Y61u1y11le1y1seu1y5a111e1g1y1n1a1s1e1s1e1c1u1l1e1h1 35
55  GGGCAATATATTTAACTCAAAAGCATTTGGACAGATAGCAGTGAAGTTCA 100
35  sPhelysVal1y5MeC1Thr1His1eu1y1s1y1s1eu1y1g1u1Se1Ty1C 52
105  CTTCAAAGTGAATAATGACAAACACATCTCAAGAAACCAAGATATCTACT 15
52  y1s1I1n1r1g1n1g1y1a1P1r1o1m1e1t1a1s1e1r1e1u1a1r1g1P1h1e1u1P1h1e1c1u1g1y1 68
155  GTCAAGACAGAGGTGTTCCTCAATGAATTCACCTAGGTTTCTCTTTAGGGT 20.
69  G1n1A1r1g1I1e1a1a1s1p1a1s1n1h1s1Th1r1P1r1o1l1y1s1e1u1e1u1g1y1M1e1t1G1u1e1G1 85
205  CACAGCAATTCGCTGATATATCATCTCCAAAAGCAACGGGGAATGAGGAAGA 25
85  u1a1s1p1a111e1g1u1a11y1r1g1n1g1u1e1G1n1Th1r1G1y1h1s1e1r1Th1r1a1 101
255  AGATGTGAATGAAGTTTATTCAGGAACAAACGGGGGGTCAATTCACAGTT 303
seq_name: gb_P19:HSSMT3C

```

seq_name: gb_pr9:HSSMT3C

seq. documentation block:	590 bp	mRNA	PRI.
LOCUS	HSSMT3C		19-MAR-1997
DEFINITION	H.sapiens mRNA for SMT3C protein.		
ACCESSION	X99586		
VERSION	X99586.1	GI:1770520	
KEYWORDS	SMT3C gene; suppressor; ubiquitin-like protein.		
SOURCE	human		

ORGANISMS

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 590)

AUTHORS

and Brahe, C.

TITLE	JOURNAL	DATE
SMT3A, a human homologue of the <i>S. cerevisiae</i> SMT3 gene, maps to chromosome 21qter and defines a novel gene family	Genomics 40 (2), 362-366 (1997)	970327050

MEDLINE 97237059

REFERENCE 1 (bases 1 to 1017)
 AUTHORS Shen, Z., Pardington-Purtymun, P.E., Comeaux, J.C., Moyzis, R.K. and Chen, D.J.
 TITLE Ubl1, a human ubiquitin-like protein associating with human RAD51/RAD52 proteins
 JOURNAL Genomics 36 (2), 271-276 (1996)
 MEDLINE 96411684
 REFERENCE 2 (bases 1 to 1017)
 AUTHORS Shen, Z.
 TITLE Direct Submission
 JOURNAL Submitted (16-OCT-1995) Zhiyuan Shen, Life Sciences Division, Los Alamos National Lab, MS M888, Los Alamos, NM 87545, USA
 FEATURES
 source location/Qualifiers
 1. 1017
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="This cDNA was cloned with the yeast two hybrid system. The protein interacts with human RAD51 and RAD52 proteins in yeast"
 CDS
 67..372
 /note="Similar to ubiquitin and to yeast Smtp (suppressor of MFP2); Method: conceptual translation supplied by author"
 /codon_start=1
 /db_xref="GI:1574948"
 /translation="MSDDEAKPSTEDLDKKEGEYIKLVIGDSSSEIHKYKMTTHLKKIKESYKROGVPMNSLRFLEGGRIADNHPKELMEEDVIEVYQOTGHSIV"
 BASE COUNT 310 a 180 c 213 g 314 t
 ORIGIN
 alignment_scores:
 Quality: 83.00 Length: 83
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-484-964-2 x HSU38784 ..
 Align seg 1/1 to: HSU38784 from: 1 to: 1017

REFERENCE 1 (bases 1 to 1187)
 AUTHORS Howe, K., Williamson, J., Boddy, N., Sheer, D., Freemont, P. and Solomon, E.
 TITLE The ubiquitin-homology gene P1C1: characterization of mouse (P1c1) and human (UBL1) genes and pseudogenes
 JOURNAL Genomics 47 (1), 92-100 (1998)
 MEDLINE 98126440
 REFERENCE 2 (bases 1 to 1187)
 AUTHORS Howe, K., Freemont, P.S. and Solomon, E.
 TITLE Direct Submission
 JOURNAL Submitted (07-NOV-1997) Medical & Molecular Genetics, Guy's Hospital, 8th floor, Guy's Tower, London SE19RT, UK
 FEATURES
 source location/Qualifiers
 1. 1187
 /organism="Mus musculus"
 /strain="ICR"
 /db_xref="taxon:10090"
 /chromosome="1"
 /map="C2-3"
 /gene="Ubl1"
 /note="P1c1"
 119..424
 /gene="Ubl1"
 /codon_start=1
 /product="ubiquitin-homology domain protein"
 /translation="MSDDEAKPSTEDLDKKEGEYIKLVIGDSSSEIHKYKMTTHLKKIKESYKROGVPMNSLRFLEGGRIADNHPKELMEEDVIEVYQOTGHSIV"
 BASE COUNT 350 a 216 c 267 g 354 t
 ORIGIN
 alignment_scores:
 Quality: 83.00 Length: 83
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-484-964-2 x AF033353 ..
 Align seg 1/1 to: AF033353 from: 1 to: 1187

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 1223)
Boddy M.N., Howe K., Etkin L.D., Solomon E. and Freemont P.S.
TITLE Plc 1, a novel ubiquitin-like protein which interacts with the pml
component of a multiprotein complex that is disrupted in acute
promyelocytic leukaemia
Oncogene 13 (5), 971-982 (1996)
JOURNAL
MEDLINE 96400311
2 (bases 1 to 1223)
REFERENCE
AUTHORS Howe K., Boddy M.N., Etkin L.D., Solomon E. and Freemont P.S.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-1996) Somatic Cell Genetics, Imperial Cancer
Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK
FEATURES
source
1. 1223
/organism="Homo sapiens"
/db_xref="taxon:9606"
/rname="placenta"
146..451
/codon_start=1
/product="ubiquitin-homology domain protein plc1"
/protein_id="AA04388.1"
/db_xref="GI:1518694"
/translation="MSDQKAPSTEDLGKKKEGYIKLVIGDSSSEIHFKVMTTHL
KKIKESYCORQGVPMNSIRFLFEGORADNTPRELQMEEDVEYVQEQGTGSHSTV"
BASE COUNT 371 a 209 c 268 g 375 t
ORIGIN

alignment_scores:
Quality: 83.00 Length: 83
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-484-964-2 x HSU61397 ..
Align seg 1/1 to: HSU61397 from: 1 to: 1223

```

19  G G G T G A T A T A T T A A C T C A A G C A T G G A C A G A T G A C A G A T T C A
|||||
200 G G T G A T A T A T T A A C T C A A G C A T G G A C A G A T G A C A G A T T C A
|||||
35  S P H E Y V A L L Y S M E T H R T H I S L E U L Y S L E U L Y S G I U S E R T Y R C
|||||
250 C T T C A A A G T G A A A T G A C A C A C A C A T C T C A A A A C T C A A A G A T C A T A C T
|||||
52  Y S G I A R G G I N G I Y A L P R O M E T A S N S E R L E U A R G P H E L U P H E G I U C L Y
|||||
300 G T C A A A G C A G G C G T T C C A A T G A A T T C A C T C A G G T T T C T C T T G A G G G T
|||||
69  G I A R G G I A L A A S P A S N I S T H R P R O L Y S G I U L E U G I M E T G I U G L G I
|||||
350 C A G A A A T T G C T G A T A T C A T C C A A A G A C T G G G A A T G A G A G A G A
|||||
85  U A S P V A L L E G I U A L Y T R G I N G I U G I N T H R G I Y G I Y H I S E R T H R Y A L
|||||
400 A G A T G T G A T G A A G T T A T T C A G A A C A A C G G G G G T C A T T C A C A G A T T
|||||
seq_name: gb_p77:BC006462

```

seq documentation block:
LOCUS BC006462 1227 bp mRNA PRI 13-APR-2001
DEFINITION Homo sapiens, ubiquitin-like 1 (sentrin), clone MGC:2095, mRNA,
complete cds.
ACCESSION BC006462
VERSION BC006462.1 GI:13623670
KEYWORDS MGC.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 1227)
Strausberg R.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIN-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.L.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
Info@bcsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kursche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Farvaneh Seedi, Jacqueline
Scheln, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 8 Row: F Column: 12.
Location/Qualifiers
1. 1227
/organism="Homo sapiens"
/db_xref="LocusID:7341"
/db_xref="taxon:9606"
/clone="MGC:2095 IMAGE:3546125"
/rname="MGC:2095 IMAGE:3546125"
/rname="MGC:2095 IMAGE:3546125"
/clone_lib="NIN_MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
132..437
/codon_start=1
/product="ubiquitin-like 1 (sentrin)"
/protein_id="AA06462.1"
/db_xref="GI:13623671"
/translation="MSDQKAPSTEDLGKKKEGYIKLVIGDSSSEIHFKVMTTHL
KKIKESYCORQGVPMNSIRFLFEGORADNTPRELQMEEDVEYVQEQGTGSHSTV"
BASE COUNT 386 a 208 c 259 g 374 t
ORIGIN

alignment_scores:
Quality: 83.00 Length: 83
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-484-964-2 x BC006462 ..
Align seg 1/1 to: BC006462 from: 1 to: 1227

```

19  G G G T G A T A T A T T A A C T C A A G C A T G G A C A G A T G A C A G A T T C A
|||||
186 G G T G A T A T A T T A A C T C A A G C A T G G A C A G A T G A C A G A T T C A
|||||
35  S P H E Y V A L L Y S M E T H R T H I S L E U L Y S L E U L Y S G I U S E R T Y R C
|||||
236 C T T C A A A G T G A A A T G A C A C A C A C A T C T C A A A A C T C A A A G A T C A T A C T
|||||
52  Y S G I A R G G I N G I Y A L P R O M E T A S N S E R L E U A R G P H E L U P H E G I U C L Y
|||||
seq_name: gb_p77:BC006462

```

286 GTCAGACACGGGTGTTCCATGATTCACCTACGCTTCTTTGAGCGT 335
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
 336 CAGAGATTCCTCATATCATCTCCAAAAGACTGGGAATGAGAGAGA 385
 85 uasPValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101
 386 AGATGTGATGTAAGTTTATCAGAACAAACGGGGGTCATTCAACAGTT 434

seq_name: gb_pat1:AR087051

seq_documentation_block:

LOCUS AR087051 1514 bp DNA PAT 07-SEP-2000
 DEFINITION Sequence 1 from patent US 5985664.
 ACCESSION AR087051
 VERSION AR087051.1 GI:10013817

KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1514)

AUTHORS Baker, B.F. and Cowser, L.M.

TITLE Antisense modulation of Sentrin expression

JOURNAL Patent: US 5985664-A 1 16-NOV-1999;

FEATURES
 source Location/Qualifiers
 1. 1514
 /organism="unknown"

BASE COUNT 462 a 256 c 314 g 482 t
 ORIGIN

alignment_scores:

Quality: 83.00 Length: 83
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AR087051 ..

Align seg 1/1 to: AR087051 from: 1 to: 1514

19 GlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGluIleH 35
 190 GGTGAATATATTAACCTCAAGTCATTGGACAGATGACAGTGAATTC 239
 35 sPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSerTyr 52
 240 CTTCAAGTGAATATGACACACATCTCAAGAACTCAAAAGATCATACT 289
 52 ysgLnaTgGlnGlyValPromeTasnSerLeuArgPheLeuPheGluGly 68
 290 GTCAGAACAGAGGTGTTCCAAATGATTCACACAGGTTCTCTTGAAGGT 339
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
 340 CAGAGATTCCTCATATCATCTCCAAAAGACTGGGAATGAGAGAGA 389
 85 uasPValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101
 390 AGATGTGATGTAAGTTTATCAGAACAAACGGGGGTCATTCAACAGTT 438

seq_name: gb_pat1:HSU83117

seq_documentation_block:

LOCUS HSU83117 1514 bp mRNA PRI 09-JAN-1997
 DEFINITION Human sentrin mRNA, complete cds.
 ACCESSION U83117.1 GI:1769601

KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 1514)
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS

Okura, T., Gong, L., Kamitani, T., Wada, T., Okura, I., Wei, C.-F.,

TITLE

Chang, H.-M., and Yen, E.T.H. Protection Against Fas/Apo-1- and Tumor Necrosis Factor-Mediated

JOURNAL

Cell Death by a Novel Protein, Sentrin

REFERENCE

J. Immunol. 157(10), 4277-4281 (1996)

AUTHORS

Okura, T., Gong, L., Kamitani, T., Wada, T., Okura, I., Wei, C.-F.,

TITLE

Chang, H.-M., and Yen, E.T.H. Direct Submission

JOURNAL

Submitted (23-DEC-1996) Division of Molecular Medicine, Department of Internal Medicine, and Cardiovascular Research Center, Institute of Molecular Medicine for the Prevention of Human Diseases, The University of Texas-Houston Health Science Center, 2121 W. Holcombe

FEATURES

Bld., Suite 900, Houston, TX 77030, USA

source

Location/Qualifiers
 1. 1514
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="placenta"

CDS

136..441
 /note="ubiquitin-like protein"
 /codon_start=1
 /product="sentrin"
 /protein_id="AAB39999.1"
 /db_xref="GI:1769602"

BASE COUNT 462 a 256 c 314 g 482 t
 ORIGIN

alignment_scores:

Quality: 83.00 Length: 83
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x HSU83117 ..

Align seg 1/1 to: HSU83117 from: 1 to: 1514

19 GlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGluIleH 35
 190 GGTGAATATATTAACCTCAAGTCATTGGACAGATGACAGTGAATTC 239
 35 sPheLysValLysMetThrThrHisLeuLysLysLeuLysGluSerTyr 52
 240 CTTCAAGTGAATATGACACACATCTCAAGAACTCAAAAGATCATACT 289
 52 ysgLnaTgGlnGlyValPromeTasnSerLeuArgPheLeuPheGluGly 68
 290 GTCAGAACAGAGGTGTTCCAAATGATTCACACAGGTTCTCTTGAAGGT 339
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
 340 CAGAGATTCCTCATATCATCTCCAAAAGACTGGGAATGAGAGAGA 389
 85 uasPValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101
 390 AGATGTGATGTAAGTTTATCAGAACAAACGGGGGTCATTCAACAGTT 438

seq_name: gb_cm:AF242526

seq_documentation_block:

LOCUS AF242526 510 bp mRNA MAM 03-AUG-2000
 DEFINITION Cervus nippon sentrin mRNA, complete cds.
 ACCESSION AF242526

KEYWORDS
 SOURCE sika deer.
 ORGANISM Cervus nippon

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae;
Cervinae; Cervinae; Cervus.

1 (bases 1 to 510)
Sun, L.G., Yu, Y.L. and Jiang, Y.
Direct Submission
Submitted (07-MAR-2000) Dept. of Immunology, Norman Bethune
University of Medical Sciences, Xin Min Street, Changchun, Jilin
130021, People's Republic of China

LOCATION/Qualifiers
1. 510
/organism="Cervus nippon"
/db_xref="taxon:9863"
/note="authority: Cervus nippon Temminck"
129. 434
/codon_start=1
/product="serpin"
/protein_id="AA097049.1"
/db_xref="GI:9664277"
/translation="MSDQAKPSTEDLDGKEGEYIKLVIGDSSSEIHKYKMTTL
KKLESYCOHGVPMNSIRLFEGRADNHFKELGMEEDYIEVYHQTGHSIV"

FEATURES

CDS
BASE COUNT 154 a 98 c 121 g 137 t
ORIGIN

alignment_scores:

Quality: 73.00 Length: 73
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AF242526 ..
Align seg 1/1 to: AF242526 from: 1 to: 510

19 GYGLUTYRILELysLeuValIleGlylnAspSerSerGluIleH 35
|||||
183 GGAGAAATATTAACTCAAGTCATTGGACAGATAGACAGATTCA 232
35 sPhelysValIysMetThrThrlsLeuLysLeuLysGluSerTyrC 52
|||||
233 CTTCAAATGAAATGACGACACATCTCAAGAACTCAAGAACTCAACT 282
52 YSGlnArGlnGlyValPrometasSerLeuArGpHeuPheGluGly 68
|||||
283 GTCAAAGACAGGAGTTCCTATGATTCACCTCAGCTTCTTTGAAGGT 332
69 GlnArGlieAlaAspAsnHisThrProLysGluLeuGluMetGluGlu 85
|||||
333 CAGAGAAATGCTGATTAATCACTCCAAAAGAACTGGGAATGAGGAAGA 382
85 uAspValIleGluValTyr 91
|||||
383 AGATGTGATTGAAGTTTAT 401

seq_name: gb_patl:AX021844

seq_documentation_block:

LOCUS AX021844 372 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 18 from Patent EP0861322.
ACCESSION AX021844
VERSION AX021844.1 GI:10045087
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 372)
AUTHORS O'Neill, R. and Palese, P.
TITLE Antiviral compounds that inhibit interaction of host cell proteins
and viral proteins required for replication
JOURNAL Patent: EP 0861322-A 18 02-SEP-1998;
MOUNT SINAI MEDICAL CENTER (US)
FEATURES Location/Qualifiers

source 1. 372
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 121 a 72 c 97 g 82 t
ORIGIN

alignment_scores:
Quality: 67.00 Length: 67
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AX021844 ..
Align seg 1/1 to: AX021844 from: 1 to: 372

19 GYGLUTYRILELysLeuValIleGlylnAspSerSerGluIleH 35
|||||
134 GGTCAAATATTAACTCAAGTCATTGGACAGATAGACAGATTCA 183
35 sPhelysValIysMetThrThrlsLeuLysLeuLysGluSerTyrC 52
|||||
184 CTTCAAAGTCAAAATGACACACATCTCAAGAACTCAAGAACTCAACT 233
52 YSGlnArGlnGlyValPrometasSerLeuArGpHeuPheGluGly 68
|||||
234 GTCAAAGACAGGAGTTCCTCAATGATTCACCTCAGCTTCTTTGAAGGT 283
69 GlnArGlieAlaAspAsnHisThrProLysGluLeuGluMetGluGlu 85
|||||
284 CAGAGAAATGCTGATTAATCACTCCAAAAGAACTGGGAATGAGGAAGA 333
85 u 85
334 A 334

seq_name: gb_htg17:AC084066

seq_documentation_block:
LOCUS AC084066 235411 bp DNA HTG 12-OCT-2000
DEFINITION Mus musculus clone RP23-321D1, *** SEQUENCING IN PROGRESS ***
ACCESSION AC084066
VERSION AC084066.1 GI:10799415
KEYWORDS HTG; HTGS_Phasel.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scturognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 235411)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 235411)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 2351294
Center clone name: RPI-23_321D1

Summary Statistics
Consensus quality: 214207 bases at least Q40
Consensus quality: 223053 bases at least Q30
Consensus quality: 225208 bases at least Q20
Estimated insert size: 200000; pulse field gel estimation


```

misc_feature      /clone.lib="RPT-11.1"
                  1..17991
                  /note="assembly-fragment:00134
                  fragment_chain:1"
misc_feature      18092..34810
                  /note="assembly-fragment:02116
                  fragment_chain:1"
misc_feature      34911..125180
                  /note="assembly-fragment:03000
                  fragment_chain:1"
misc_feature      125281..129032
                  /note="assembly-fragment:01240
                  fragment_chain:1"
misc_feature      129133..132599
                  /note="assembly-fragment:00856
                  fragment_chain:1"
misc_feature      132700..141603
                  /note="assembly-fragment:00528"
misc_feature      141704..167440
                  /note="assembly-fragment:02701"
BASE COUNT      47457 a 37675 c 35802 g 45906 t      600 others
ORIGIN
alignment_scores:
    Quality:      43.00      Length:      43
    Ratio:         1.000      Gaps:         0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-484-964-2 x ALJ359981/rev ..

Align seg 1/1 to reverse of: ALJ359981 from: 1 to: 167440

19 GYGIUTYRIleYsLeuYsValIleGIgInaSPserSerGIuIleH1 35
|||||
154373 GGTGAAATATATTAACCTCAAACTCATTTGACAGCATGACAGTTCATCA 154324
35 sPhelYsValYsMeLthrThrHisLeuYsLysLeuYsGluSerTYRC 52
|||||
154323 CTTCAAAGTGAATATGACAAACATCTCAAGAACTCAAGAAATCATACT 154274
52 YSGIARGInGInGlyValIPrometAsnSer 61
|||||
154273 GTCCAAAGACAGGGCGCTCCAAATGAATTCATCA 154245

seq_name: gb_pf5:ALJ35926

seq_documentation_block:
LOCUS      ALJ35926      158519 bp      DNA      30-NOV-2000
DEFINITION      Human DNA sequence from clone RP11-375F2 on chromosome 1 Contains a
                  pseudogene similar to UBL1 (ubiquitin-like 1 (sentrin)), a
                  pseudogene similar to ribosomal protein L29, ESTs, STSS and GSSs,
                  complete sequence.
ACCESSION      ALJ35926
VERSION        ALJ35926.12      GI:9801286
KEYWORDS       HTG.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 158519)
AUTHORS       Chapman,J.
TITLE         Direct Submission
JOURNAL       Submitted (03-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
COMMENT
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
validation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with

```

FEATURES
SOURCE

Only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WormPEP; Information on the WormPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP11-375F2 is from the library RCP1-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://pacpac.med.buffalo.edu/VECTOR:PBACE3.6>

IMPORTANT: This sequence is not the entire insert of clone RP11-375F2. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-375F2 is at 158519 in this sequence. The true right end of clone RP1-10C16 is at 100 in this sequence.

```

RRES
source
    1. 158519
       /organism="Homo sapiens"
       /db_xref="taxon:9606"
       /chromosome="11"
       /clone="RP11-375F2"
       /clone_11b="RCP1-11.2"
       460. 675
       /note="MTL1L repeat: matches 306. 514 of consensus"
repeat_region
    1465. 1514
       /note="75 copies 2 mer to 72% conserved"
       1775. 1824
       /note="MIR repeat: matches 85. 135 of consensus"
       3415. 3734
       /note="L2 repeat: matches 2373. 2707 of consensus"
       4294. 5519
       /note="L1MB7 repeat: matches 4938. 6173 of consensus"
       5355. 5640
       /note="L1MB5 repeat: matches 5322. 5428 of consensus"
       5652. 5945
       /note="L1MB4 repeat: matches 5322. 5428 of consensus"
       5960. 5995
       /note="AluSq repeat: matches 12. 309 of consensus"
       /note="9 copies 4 mer aagg 88% conserved"
       5997. 6055
       /note="L1MB5 repeat: matches 5413. 5471 of consensus"
       6061. 6158
       /note="AluY repeat: matches 214. 311 of consensus"
       6160. 6855
       /note="L1MB5 repeat: matches 5460. 6168 of consensus"
       6919. 7101
       /note="L2 repeat: matches 2206. 2403 of consensus"
       7099. 7228
       /note="L2 repeat: matches 2620. 2750 of consensus"
       7828. 8069
       /note="MEB46a repeat: matches 2. 235 of consensus"
       8185. 8469
       /note="AluXs repeat: matches 1. 285 of consensus"
       8551. 8624
       /note="MIR repeat: matches 81. 153 of consensus"
       8951. 9347
       /note="match: SIS: Em:B40536"
       10287. 10464
       /note="MIR repeat: matches 28. 215 of consensus"
       10568. 11321
       /note="L1PA8 repeat: matches 5371. 6158 of consensus"
       11337. 11671
       /note="MT1D1 repeat: matches 13. 381 of consensus"

```

```

repeat_region 11672..11971
/misc_feature 11855..12218
repeat_region 11972..12090
repeat_region 12106..12403
repeat_region 12419..13151
/misc_feature 13091..13572
repeat_region 13164..13253
repeat_region 13254..13403
repeat_region 13267..13410
repeat_region 13654..13742
repeat_region 14045..14226
/misc_feature 14901..15366
repeat_region 15371..15463
repeat_region 16536..16633
repeat_region 16839..17215
repeat_region 17225..17961
repeat_region 17962..19671
repeat_region 19672..20271
repeat_region 20277..20462
repeat_region 20479..21023
repeat_region 21083..22136
repeat_region 22104..24249
repeat_region 24175..24890
repeat_region 24896..25532
repeat_region 25865..25900
repeat_region 25903..26184
/misc_feature 27110..27647
repeat_region 27233..27346
repeat_region 27347..27437
repeat_region 27438..27487
repeat_region 27552..27933
repeat_region 27982..28315
repeat_region 28256..28552
repeat_region 28629..28938
repeat_region 29045..29218
repeat_region 30188..30300

```

```

repeat_region 30761..30877
repeat_region 31253..31494
gene 31722..32025
CDS 31722..32025
/gene="BA375F2.1"
/gene="BA375F2.1"
/gene="BA375F2.1" (similar to UBL1 (ubiquitin-like 1
(sentrin)))
match: cDNAs: Em:U72722
match: ESTs: Em:A119122 Em:A1248769
match: proteins: Tr:Q9PT08 Tr:Q23759 Sw:P55853 Tr:Q92172
Sw:Q93068 Sw:P55854 Sw:P55855 Sw:O13351 Sw:Q12306
Tr:Q97102 Tr:Q57686"
/codon_start=1
/pseudo
evidence=not_experimental
complement(32129..32630)
/misc_feature 32225..32619
/misc_feature 32225..32619
/misc_feature 32225..32619
complement(32404..32630)
repeat_region 33174..33222
repeat_region 33291..33725
repeat_region 35465..35500

```

```

alignment_scores:
Quality: 42.00 Length: 42
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-484-964-2 x AL135926

```

```

Align seg 1/1 to: AL135926 from: 1 to: 158519

```

```

23 LysLeuLysValIleGlyGlnAspSerSerGluIleHisPheLysValIly 39
|||||
31788 AAATCAAAATCATTTGGACAGGATACGACGAGATCTTCAAACTGAA 31837
|||||
39 SMeTThrHisLeuLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 56
|||||
31838 AATGACAAACATCTCAGAACTCAAGAAATCATCTGCTCAAGGACAG 31887
|||||
56 LysAlaPrometAsnSerLeuArgPhe 64
|||||
31888 GAGTTCATGAAATTCACCTCAGATT 31913

```

OM of: US-09-484-964-2 to: N_Geneseq_0601.* out_format: pfs

Date: Aug 15, 2001 6:43 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed+.p2n.model -DEV=xlp
-O=/cgn2.1/USPPO.spool/US09484964/runtat_14082001_111933_27731/app-query.fasta_1.158
-DB=N_Geneseq_0601 -OPMT=fastap -SUFFIX=oli.rng -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FEADOP=6.000 -DELEXT=7.000 -XGAPOP=60.000 -XGAPEXT=60.000
-FEADOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oli1go
-TRANS=human4.0.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=quality
-THR.MIN=20 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09484964.ecgn1_1.190 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-484-964-2
Query length: 101
Database: N_Geneseq_0601.*
Database sequences: 730101
Database length: 313950809
Search time (sec): 128.620000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation	..	1465
/SID1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV34564				101.00	1905.04	3.8e-98	1465
/SID1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV34564				83.00	1563.51	4.1e-79	1196
/SID1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV59722				83.00	1563.37	4.1e-79	1220
/SID1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV59557				83.00	1561.83	5.0e-79	1514
/SID1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV35859				71.00	1339.61	1.2e-66	616
/SID1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA16215				67.00	1267.00	1.3e-62	372
/SID1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV63339				67.00	1267.00	1.3e-62	372
/SID1/gcgdata/geneseq/geneseq/NA2001.DAT:AAV59399				65.00	1227.94	2.0e-60	425
/SID1/gcgdata/geneseq/geneseq/NA1996.DAT:AAV03735				32.00	600.84	1.7e-25	335
/SID1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA3182							

seq_name: /SID1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV34564

seq_documentation_block:

ID AAV34564 standard; DNA: 1465 BP.

XX AAV34564:

DT 11-SEP-1998 (first entry)

XX Homo sapiens sentrin-1 gene.

DE sentrin-1; protection; tumour necrosis factor; apoptosis;

KW Fas/Apo-induced; tumour cell death; induction; tumour aggressiveness;

KW detection; determination; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 88..393

XX FT /*tag= a

XX FT /product= sentrin-1 polypeptide

XX PN WO9820038-A1.

XX PD 14-MAY-1998.

XX PF 05-NOV-1997; 97WO-US20344.

XX PR 05-NOV-1996; 96US-0030302.

XX (TEXA) UNIV TEXAS SYSTEM.

XX yeh ETH;

XX WPI: 1998-286868/25.

XX P-PSDB: AAW60079.

XX New isolated sentrin polypeptide(s) - which inhibit TNF receptor or

XX Fas/Apo-induced apoptosis, used to develop products for inducing

XX cell death in tumours

XX Claim 10; Page 76-77; 120pp: English.

XX The sequence is that encoding the sentrin-1 polypeptide.

XX Sentrin polypeptides have the ability to protect or guard

XX cells from tumour necrosis factor (TNF) or Fas/Apo induced

XX cell death (apoptosis). Inhibitors of the sentrin polypeptides,

XX e.g. antibodies, can be used for inducing cell death,

XX particularly in tumours. The products can also be used for

XX determining the aggressiveness of a tumour and for detection and

XX isolation of products. The sentrin polypeptide can also be used to

XX detect a ubiquitin conjugating enzyme polypeptide or PML polypeptide.

SO Sequence 1465 BP; 455 A; 248 C; 292 G; 470 T; 0 other;

alignment_scores:

Quality: 101.00 Length: 101

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AAV34564 ..

Align seg 1/1 to: AAV34564 from: 1 to: 1465

1 MetSerAspGlnGluAlaLysProSerThrGlnAspLeuGlyAspLysLys 17

88 ATGTGTGACCGAGGAGGCAAACTTCAAACTGAGACCTGGGGCATAGAA 137

17 scInGlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerLut 34

138 GCAAGTGAATATATTAACTCAAAAGTCATGACAGGATAGCGTGAGA 187

34 LeuIshPheLysValLysMetThrTrpHisLeuLysLysLeuLysGluSer 50

188 TTCACCTCAAAAGTGAATGACACACATCTCAAGAACTCAAAAGATCA 237

51 TYRCYSGlnArgGlnGlyValProMetAsnSerLeuArgPheLeuPheG1 67

238 TACTGTCAAAAGACAGGGTGTTCATGAAATTCACACAGTTCCTTGA 287

67 uGlyGlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlu 84

288 GGGTCAGAGAAATGGCTGATATATCATCTCCAAAGACATGGGAATGGG 337

84 IncLuspyValIleGlyValIleGlyGlnGlnGlnThrGlyHisSerThr 100

338 AAGAAGATGTGATTGAATTTATCAAGAACAAACGGGGGTCATTCAACA 387

seq_name: /SID1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV59722

seq_documentation_block:

ID AAV59722 standard; DNA: 1196 BP.

XX AAV59722;

XX DT 19-JAN-1999 (first entry)

XX DE Human secreted protein gene 47 clone HOGAV75.
 XX XX
 KW Human: secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 XX WO9839448-A2.
 PD 11-SEP-1998.
 XX
 PF 06-MAR-1998; 98WO-US04493.
 XX
 PR 02-OCT-1997; 97US-0061060.
 PR 07-MAR-1997; 97US-0038621.
 PR 07-MAR-1997; 97US-0040161.
 PR 07-MAR-1997; 97US-0040162.
 PR 07-MAR-1997; 97US-0040163.
 PR 07-MAR-1997; 97US-0040333.
 PR 07-MAR-1997; 97US-0040334.
 PR 07-MAR-1997; 97US-0040336.
 PR 11-APR-1997; 97US-0043311.
 PR 11-APR-1997; 97US-0043312.
 PR 11-APR-1997; 97US-0043313.
 PR 11-APR-1997; 97US-0043314.
 PR 11-APR-1997; 97US-0043568.
 PR 11-APR-1997; 97US-0043569.
 PR 11-APR-1997; 97US-0043576.
 PR 11-APR-1997; 97US-0043578.
 PR 11-APR-1997; 97US-0043580.
 PR 11-APR-1997; 97US-0043589.
 PR 11-APR-1997; 97US-0043670.
 PR 11-APR-1997; 97US-0043671.
 PR 11-APR-1997; 97US-0043672.
 PR 11-APR-1997; 97US-0043674.
 PR 23-MAY-1997; 97US-0047492.
 PR 23-MAY-1997; 97US-0047500.
 PR 23-MAY-1997; 97US-0047501.
 PR 23-MAY-1997; 97US-0047502.
 PR 23-MAY-1997; 97US-0047503.
 PR 23-MAY-1997; 97US-0047581.
 PR 23-MAY-1997; 97US-0047582.
 PR 23-MAY-1997; 97US-0047583.
 PR 23-MAY-1997; 97US-0047584.
 PR 23-MAY-1997; 97US-0047585.
 PR 23-MAY-1997; 97US-0047586.
 PR 23-MAY-1997; 97US-0047587.
 PR 23-MAY-1997; 97US-0047588.
 PR 23-MAY-1997; 97US-0047589.
 PR 23-MAY-1997; 97US-0047590.
 PR 23-MAY-1997; 97US-0047592.
 PR 23-MAY-1997; 97US-0047593.
 PR 23-MAY-1997; 97US-0047594.
 PR 23-MAY-1997; 97US-0047595.
 PR 23-MAY-1997; 97US-0047596.
 PR 23-MAY-1997; 97US-0047597.
 PR 23-MAY-1997; 97US-0047598.
 PR 23-MAY-1997; 97US-0047599.
 PR 23-MAY-1997; 97US-0047600.
 PR 23-MAY-1997; 97US-0047601.
 PR 23-MAY-1997; 97US-0047610.
 PR 23-MAY-1997; 97US-0047613.
 PR 23-MAY-1997; 97US-0047614.
 PR 23-MAY-1997; 97US-0047615.
 PR 23-MAY-1997; 97US-0047617.

PR 23-MAY-1997; 97US-0047618.
 PR 23-MAY-1997; 97US-0047632.
 PR 23-MAY-1997; 97US-0047633.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048974.
 PR 13-JUN-1997; 97US-0048974.
 PR 08-JUL-1997; 97US-0049610.
 PR 16-JUL-1997; 97US-0051926.
 PR 18-AUG-1997; 97US-0052874.
 PR 22-AUG-1997; 97US-0055724.
 PR 22-AUG-1997; 97US-0056630.
 PR 22-AUG-1997; 97US-0056631.
 PR 22-AUG-1997; 97US-0056632.
 PR 22-AUG-1997; 97US-0056636.
 PR 22-AUG-1997; 97US-0056637.
 PR 22-AUG-1997; 97US-0056662.
 PR 22-AUG-1997; 97US-0056664.
 PR 22-AUG-1997; 97US-0056665.
 PR 22-AUG-1997; 97US-0056662.
 PR 22-AUG-1997; 97US-0056664.
 PR 22-AUG-1997; 97US-0056672.
 PR 22-AUG-1997; 97US-0056874.
 PR 22-AUG-1997; 97US-0056875.
 PR 22-AUG-1997; 97US-0056876.
 PR 22-AUG-1997; 97US-0056877.
 PR 22-AUG-1997; 97US-0056877.
 PR 22-AUG-1997; 97US-0056878.
 PR 22-AUG-1997; 97US-0056879.
 PR 22-AUG-1997; 97US-0056880.
 PR 22-AUG-1997; 97US-0056881.
 PR 22-AUG-1997; 97US-0056882.
 PR 22-AUG-1997; 97US-0056884.
 PR 22-AUG-1997; 97US-0056886.
 PR 22-AUG-1997; 97US-0056887.
 PR 22-AUG-1997; 97US-0056888.
 PR 22-AUG-1997; 97US-0056889.
 PR 22-AUG-1997; 97US-0056892.
 PR 22-AUG-1997; 97US-0056893.
 PR 22-AUG-1997; 97US-0056894.
 PR 22-AUG-1997; 97US-0056903.
 PR 22-AUG-1997; 97US-0056908.
 PR 22-AUG-1997; 97US-0056909.
 PR 22-AUG-1997; 97US-0056910.
 PR 22-AUG-1997; 97US-0056911.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057669.
 PR 05-SEP-1997; 97US-0057761.
 PR 12-SEP-1997; 97US-0058785.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Bedharik DP, Brewer LA, Carter KC, Duan R, Edner R, Endress GA;
 PI Feng P, Ferris AM, Fischer CL, Florence KA, Greene JM, Hu JS;
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 XX
 DR WPI: 1998-506364/43.
 DR P-PSDB: AAW74938.
 XX
 PT New isolated human genes and the secreted polypeptide(s) they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1; Page 449; 721pp; English.
 XX
 CC This sequence represents a nucleic acid molecule designated Gene 47 from
 CC the human cDNA clone HOGAV75 (deposited as clone ATCC 97897 and ATCC
 CC 209043) which encodes a secreted human protein. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
 CC the fused protein as compared to the human protein only.
 CC The invention relates to 186 novel genes and their fragments (nucleic
 CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological

CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 186
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAV59511 for described uses).

XX Sequence 1196 BP; 360 A; 212 C; 259 G; 364 T; 1 other;

alignment_scores: Quality: 83.00 Length: 83

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AAV59722 ..

Align seg 1/1 to: AAV59722 from: 1 to: 1196

19 GYGIUITYRIELVELYELYSVAIIIEGLYGLINAsPserSergIuIIeH 35
 |||||||
 188 GGTGATATATTAACTCAAGCTCTTTGACAGATACAGTACAGATTCA 237
 35 sPhelysValysMetThrHrHisLeuLysLysLeuLysGluSerTYRC 52
 |||||||
 238 CTTCAAGTGAANAATGACACACATCTCAGAACTCAAGATCATCTACT 287
 52 YGGLARGInGInGlyValPrometAsnSerLeuArgPheLeuPheGluGly 68
 |||||||
 288 GTCAAGAGAGAGGGGTGTCATCAATTCATCTAGTTCTCTTAAAGGT 337
 69 GlnArgIleAlaASPAsnHisThrProLysGluLeuGlyMetGluGlu 85
 |||||||
 338 CAGAGATTGCTGATATCATCTCCAAAGAACTGGCAATGAGAGAGA 387
 85 uAspValIIEgluValTYrGInGluGInThrGlyGlnHisSerThrVal 101
 |||||||
 388 AGATGATGATGAAGTTTATCAGAGAACAAAGGGGGTCTTCAACAGTT 436
 seq_name: /SID1/cgdata/geneseq/geneseq/NA1998.DAT:AAV59557
 seq_documentation_block:
 ID AAV59557 standard; DNA; 1220 BP.
 AC AAV59557;
 XX
 XX
 DT 06-JAN-1999 (first entry)
 XX
 XX
 DE Human secreted protein gene 47 clone HOGANV75.
 XX
 XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 XX
 PN WO9839448-A2.
 XX
 PD 11-SEP-1998.
 XX
 XX
 PF 06-MAR-1998; 98WO-US04493.
 XX
 XX
 PR 02-OCT-1997; 97US-0061060.
 PR 07-MAR-1997; 97US-0038621.
 PR 07-MAR-1997; 97US-0040161.
 PR 07-MAR-1997; 97US-0040162.
 PR 07-MAR-1997; 97US-0040163.
 PR 07-MAR-1997; 97US-0040333.

PR 07-MAR-1997; 97US-0040334.
 PR 07-MAR-1997; 97US-0040336.
 PR 07-MAR-1997; 97US-0040626.
 PR 11-APR-1997; 97US-0043311.
 PR 11-APR-1997; 97US-0043312.
 PR 11-APR-1997; 97US-0043313.
 PR 11-APR-1997; 97US-0043314.
 PR 11-APR-1997; 97US-0043568.
 PR 11-APR-1997; 97US-0043569.
 PR 11-APR-1997; 97US-0043576.
 PR 11-APR-1997; 97US-0043578.
 PR 11-APR-1997; 97US-0043580.
 PR 11-APR-1997; 97US-0043669.
 PR 11-APR-1997; 97US-0043670.
 PR 11-APR-1997; 97US-0043671.
 PR 11-APR-1997; 97US-0043672.
 PR 11-APR-1997; 97US-0043674.
 PR 23-MAY-1997; 97US-0047492.
 PR 23-MAY-1997; 97US-0047501.
 PR 23-MAY-1997; 97US-0047502.
 PR 23-MAY-1997; 97US-0047503.
 PR 23-MAY-1997; 97US-0047504.
 PR 23-MAY-1997; 97US-0047581.
 PR 23-MAY-1997; 97US-0047582.
 PR 23-MAY-1997; 97US-0047583.
 PR 23-MAY-1997; 97US-0047584.
 PR 23-MAY-1997; 97US-0047585.
 PR 23-MAY-1997; 97US-0047586.
 PR 23-MAY-1997; 97US-0047587.
 PR 23-MAY-1997; 97US-0047588.
 PR 23-MAY-1997; 97US-0047589.
 PR 23-MAY-1997; 97US-0047590.
 PR 23-MAY-1997; 97US-0047592.
 PR 23-MAY-1997; 97US-0047593.
 PR 23-MAY-1997; 97US-0047594.
 PR 23-MAY-1997; 97US-0047595.
 PR 23-MAY-1997; 97US-0047596.
 PR 23-MAY-1997; 97US-0047597.
 PR 23-MAY-1997; 97US-0047598.
 PR 23-MAY-1997; 97US-0047599.
 PR 23-MAY-1997; 97US-0047601.
 PR 23-MAY-1997; 97US-0047602.
 PR 23-MAY-1997; 97US-0047612.
 PR 23-MAY-1997; 97US-0047613.
 PR 23-MAY-1997; 97US-0047614.
 PR 23-MAY-1997; 97US-0047615.
 PR 23-MAY-1997; 97US-0047617.
 PR 23-MAY-1997; 97US-0047618.
 PR 23-MAY-1997; 97US-0047632.
 PR 23-MAY-1997; 97US-0047633.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048974.
 PR 13-JUN-1997; 97US-0049610.
 PR 08-JUL-1997; 97US-0051926.
 PR 16-JUL-1997; 97US-0052874.
 PR 18-AUG-1997; 97US-0055724.
 PR 22-AUG-1997; 97US-0056630.
 PR 22-AUG-1997; 97US-0056631.
 PR 22-AUG-1997; 97US-0056632.
 PR 22-AUG-1997; 97US-0056636.
 PR 22-AUG-1997; 97US-0056637.
 PR 22-AUG-1997; 97US-0056662.
 PR 22-AUG-1997; 97US-0056664.
 PR 22-AUG-1997; 97US-0056845.
 PR 22-AUG-1997; 97US-0056862.
 PR 22-AUG-1997; 97US-0056864.
 PR 22-AUG-1997; 97US-0056872.
 PR 22-AUG-1997; 97US-0056874.
 PR 22-AUG-1997; 97US-0056875.
 PR 22-AUG-1997; 97US-0056876.
 PR 22-AUG-1997; 97US-0056877.
 PR 22-AUG-1997; 97US-0056878.
 PR 22-AUG-1997; 97US-0056879.

US-09-484-964-2 x AAZ35859

Align seg 1/1 to: AAZ35859 from: 1 to: 1514

```

19 GYGLUTYRILEYLSLEULYSVALILEGLINASPSESGIULEHI 35
190 GGTGAATATTTAACTCAAAAGTCATGGACAGATAGCAGATTC 239
35 SPHELYSVALYSMETHTHRHISLEULYSLEULYSGLUSERTYRC 52
240 CTTCAAAGTGAATAATGACACATCTCAAGAACTCAAAAGTCACTACT 289
52 YSGINARGINGLYVALIPROMETASNSERLEUARGPHELEUPHEGLUGLY 68
290 GTCAAAGACAGGGGTTCCTCAATGATCACTACAGTTCTCTTGAGGCT 339
69 GLNARGTLEALASPASNHSRHPROLYSGULLEUGLYMETGLUGL 85
340 CAGAGAATTCCTGATTAATCATCTCCAAAAGAACTGGGAATGGAGAGA 389
85 UASPVALLIEGLUVALTYRGINGLUGINTHNGLYGHISETHTRYAL 101
390 AGAGTGTATGTAAGTTATTCAGGACAAACGGGGTCATTCAACAGTT 438
seq_name: /SIDSI/gcgcdata/geneseq/geneseqn/NA2000.DAT:AAA16215
seq_documentation_block:
ID AAA16215 standard; DNA; 616 BP.
XX
AC AAA16215;
XX
DT 14-JUN-2000 (first entry)
XX
DE Human colon cancer differentially expressed nucleotide sequence #220.
XX
KW Colon cancer; detect; differential expression; human; treatment;
KM detect mutation; non-invasive diagnostic method; ds.
XX
OS Homo sapiens.
XX
PN M0200012702-A2.
XX
PD 09-MAR-2000.
XX
PE 30-AUG-1999; 99WO-US19424.
XX
PR 31-AUG-1998; 98US-0098639.
PR 27-JAN-1999; 99US-0117393.
XX
PA (FARB ) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
PI Carino TJ, Dwaivedl P, Ford DM, Lewis ME, Molino GA, Monahan JE;
PI Schlegel R;
XX
WPI: 2000-256641/22.
XX
PT Novel nucleic acids and proteins for identifying therapeutic agents
PT useful for treating and diagnosing cancer, especially colon cancer
XX
PS Claim 16; Page 211-212; 345pp; English.
XX

```

This sequence represents a human nucleotide sequence which is differentially expressed in colon cancer cells compared to the expression levels in normal cells. The nucleotide sequence can be used as a source of primers and probes. The nucleotide sequence is useful for determining the phenotype of a cell by detecting the differential expression of the sequence relative to a normal cell. The probes derived from the sequence can also be used to determine the phenotype of cells in a sample. Probes and antibodies which hybridize to the nucleotide sequence can also be used to determine the phenotype of a cell. The primers are useful for detecting a mutation in a test nucleotide sequence and also for detecting cancer, preferably colon cancer. Antibodies against the protein encoded

CC by the nucleotide sequence can also be used in a method to detect colon cancer. The diagnostic method is non-invasive and accurate for diagnosing colon cancer at an early stage.

SO Sequence 616 BP; 181 A; 109 C; 149 G; 170 T; 7 other;

alignment_scores:

Quality:	Length:
Ratio: 1.000	Gaps: 0
Percent Similarity: 100.000	Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AAA16215

Align seg 1/1 to: AAA16215 from: 1 to: 616

```

19 GYGLUTYRILEYLSLEULYSVALILEGLINASPSESGIULEHI 35
181 GGTGAATATTTAACTCAAAAGTCATGGACAGATAGCAGATTC 230
35 SPHELYSVALYSMETHTHRHISLEULYSLEULYSGLUSERTYRC 52
231 CTTCAAAGTGAATAATGACACATCTCAAGAACTCAAAAGTCACTACT 280
52 YSGINARGINGLYVALIPROMETASNSERLEUARGPHELEUPHEGLUGLY 68
281 GTCAAAGACAGGGGTTCCTCAATGATCACTACAGTTCTCTTGAGGCT 330
69 GLNARGTLEALASPASNHSRHPROLYSGULLEUGLYMETGLUGL 85
331 CAGAGAATTCCTGATTAATCATCTCCAAAAGAACTGGGAATGGAGAGA 380
85 UASPVALLIEGLU 89
381 AGAGTGTATGTAAGTTATTCAGGACAAACGGGGTCATTCAACAGTT 393

```

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn/NA1997.DAT:AA63339

seq_documentation_block:

ID AAT63339 standard; CDNA; 372 BP.

```

XX
AC AAT63339;
XX
DT 17-AUG-1997 (first entry)
XX
DE Human host cell protein NPI-6 partial cDNA clone.
XX
KW Host cell protein; NPI-6; nucleoprotein interactor 6;
KM Influenza virus; replication; antiviral; virucide; ss.
XX
OS Homo sapiens.
XX
PN M09712967-A1.
XX
PD 10-APR-1997.
XX
PE 06-OCT-1995; 95WO-US13044.
XX
PR 06-OCT-1995; 95WO-US13044.
XX
PA (MOUN ) MOUNT SINAI MEDICAL CENT.
XX
PI Onelli R, Palese P;
XX
WPI: 1997-226211/20.
XX
PT New isolated DNA which encodes viral interacting proteins - used in
PT assays to isolate products for inhibiting viral protein binding
PT which is required for infection, replication, assembly or release
XX
PS Disclosure; Fig 11; 98pp; English.
XX

```

CC cDNA clones (AAT63335-39) comprise partial sequences for human
 CC nucleoprotein interactor proteins NP1-2 to NP1-6, respectively.
 CC These are host cell proteins which interact with influenza virus
 CC nucleoprotein (NP) and which may be accessory proteins required for
 CC influenza virus replication. NP1-6 was identified by interactive
 CC protein sequence. NP1 sequences were isolated by interactive trap
 CC selection using LexA-NP as bait and yeast transformed with an HeLa
 CC cDNA library. NP1-1 (see also AAT63334) and NP1-2 to NP1-6 cDNAs were
 CC cloned and analysed. The clones can be used to produce human NP1
 CC proteins for use in identifying cpds. that inhibit viral
 CC replication.

SO Sequence 372 BP; 121 A; 72 C; 97 G; 82 T; 0 other;

alignment_scores:
 Quality: 67.00 Length: 67
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AAT63339

Align seg 1/1 to: AAT63339 from: 1 to: 372

```

19 G1G1G1T1T1T1L1E1L1S1L1E1L1S1V1A1I1E1G1L1N1A1S1P1S1E1R1G1U1I1E1H1 35
   |||||||
134 GGTGAATATATTAACCTCAAGTCATGTGACAGATGACAGATGATCA 183
   |||||||
35 SPHELYSVAL1YSMETHTH1RH1S1L1E1L1S1L1E1L1S1L1E1L1S1G1U1S1E1R1T1Y1C 52
   |||||||
184 CTTCAAGTGAATAATGACACACATCTCAAGAACTCAAGAAATCATCTACT 233
   |||||||
52 YSG1N1A1R1G1N1G1Y1A1L1P1R1O1M1E1A1S1E1R1L1E1U1A1R1G1P1H1E1U1P1H1E1G1U1G1 68
   |||||||
234 GTC1A1A1G1A1C1A1G1G1G1T1T1C1C1A1T1G1A1T1T1C1T1C1T1T1G1A1G1G1T 283
   |||||||
69 G1N1A1R1G1I1E1A1A1S1P1A1N1H1S1T1H1R1P1R1O1L1Y1S1G1U1E1U1G1Y1M1E1T1G1U1G1 85
   |||||||
284 CAGAG1A1T1T1G1T1G1A1T1A1T1C1A1T1C1C1A1A1A1G1A1C1T1G1G1A1T1G1A1G1A1G1A 333
   |||||||
85 u 85
   |
334 A 334

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT:AAT63399

seq_documentation_block:

ID AAF59399 standard; DNA; 372 BP.

AC AAF59399;

DT 02-MAY-2001 (first entry)

DE Human host cell protein NP1-6 partial nucleotide sequence SEQ ID NO:11.

KW Identification: antiviral; viral protein; viral replication; NP;

KW viral infection; nucleoprotein; ds.

OS Homo sapiens.

PN WO200111335-A2.

PD 15-FEB-2001.

PF 11-AUG-2000; 2000WO-US22257.

PR 11-AUG-1999; 99US-0148263.

PA (MOON) MOUNT SINAI SCHOOL MEDICINE.

PI O'Neill R, Hartly R, Palese PW;

XX

DR WPI; 2001-168816/17.

XX Identifying a substance that inhibits the interaction between a viral
 PT protein and a host cell protein, useful for the discovery of new
 PT antiviral compounds -

PS Disclosure; Fig 11; 147pp; English.

XX The present invention describes a method (M1) for identifying a
 CC substance that inhibits the interaction of a viral protein (VP) with a
 CC host cell protein (HP). The method comprises: (a) contacting HP with VP
 CC in the presence of a test substance; and (b) detecting complex formation,
 CC where the ability of the test substance to inhibit HP/VP interaction is
 CC indicated by a decrease in complex formation. The antiviral compounds
 CC that inhibit the interaction between a host protein (NS1-BP or NP1-1)
 CC and a viral protein (NS1) are useful for treating or inhibiting viral
 CC infection, preferably influenza and rhinovirus infection, in humans.
 CC Antiviral compounds include peptides and antibodies. In particular
 CC compositions comprising a polypeptide containing an amino acid sequence
 CC corresponding to the NP-NLS domain of the influenza virus NP protein,
 CC which inhibits the specific interaction of the NP1-1 protein with the
 CC influenza virus NP protein are useful for treating or inhibiting
 CC human host cell protein NP1-6 partial nucleotide sequence, which is used
 CC in the exemplification of the present invention.

SO Sequence 372 BP; 121 A; 72 C; 97 G; 82 T; 0 other;

alignment_scores:
 Quality: 67.00 Length: 67
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AAF59399

Align seg 1/1 to: AAF59399 from: 1 to: 372

```

19 G1G1G1T1T1T1L1E1L1S1L1E1L1S1V1A1I1E1G1L1N1A1S1P1S1E1R1G1U1I1E1H1 35
   |||||||
134 GGTGAATATATTAACCTCAAGTCATGTGACAGATGACAGATGATCA 183
   |||||||
35 SPHELYSVAL1YSMETHTH1RH1S1L1E1L1S1L1E1L1S1L1E1L1S1G1U1S1E1R1T1Y1C 52
   |||||||
184 CTTCAAGTGAATAATGACACACATCTCAAGAACTCAAGAAATCATCTACT 233
   |||||||
52 YSG1N1A1R1G1N1G1Y1A1L1P1R1O1M1E1A1S1E1R1L1E1U1A1R1G1P1H1E1U1P1H1E1G1U1G1 68
   |||||||
234 GTC1A1A1G1A1C1A1G1G1G1T1T1C1C1A1T1G1A1T1T1C1T1C1T1T1G1A1G1G1T 283
   |||||||
69 G1N1A1R1G1I1E1A1A1S1P1A1N1H1S1T1H1R1P1R1O1L1Y1S1G1U1E1U1G1Y1M1E1T1G1U1G1 85
   |||||||
284 CAGAG1A1T1T1G1T1G1A1T1A1T1C1A1T1C1C1A1A1A1G1A1C1T1G1G1A1T1G1A1G1A1G1A 333
   |||||||
85 u 85
   |
334 A 334

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT03735

seq_documentation_block:

ID AAT03735 standard; cDNA; 425 BP.

AC AAT03735;

DT 26-MAR-1996 (first entry)

DE TNF-R p55IC/Fas-IC-binding protein D11 cDNA clone.

KW Tumour necrosis factor receptor; TNF-R; p55IC; Fas-IC;

KW intracellular domain binding protein; ss.

XX

OS Homo sapiens.
 XX
 XX MO9531544-A1.
 XX
 PD 23-NOV-1995.
 XX
 PF 11-MAY-1995; 95WO-US05854.
 XX
 PR 02-OCT-1994; 94IL-0111125.
 PR 11-MAY-1994; 94IL-0109632.
 XX
 PA (WEIN/) WEINMURZEL H.
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Boldin M, Mett I, Varfolomeev E, Wallach D;
 DR WPI; 1996-010930/01.
 XX
 XX TNF-NGF receptor superfamily intracellular domain-binding proteins -
 PT useful for modulating receptor function, e.g. for treating tumours
 PT or HIV-infected cells
 XX
 PS Claim 18; Fig 12; 96pp; English.
 XX
 CC CDNA clone DD11 (AAT03735) was isolated by screening a human HeLa
 CC library using the human tumour necrosis factor receptor (TNF-R)
 CC p55 death domain (p55DD) as 'bait'. The full length of the DD11
 CC transcript is approx 1.2 kb. The protein encoded by CDNA DD11
 CC interacts strongly with p55DD (amino acids 326-414) and also with
 CC human and mouse FAS intracellular domain. Such proteins may
 CC be used modulate TNF-R function.
 XX
 SQ Sequence 425 BP; 135 A; 80 C; 115 G; 95 T; 0 other;

alignment_scores:
 Quality: 65.00 Length: 65
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-484-964-2 x AAT03735 ..

Align seg 1/1 to: AAT03735 from: 1 to: 425

19 G1yG1uTyr11e11e1y1s1e1u1y1s1a11e1g1y1G1n1A1s1p1e1s1e1r1u1l1e1n1 35
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 173 GAGCAATATATATAAAGTCAAAAGTCATGACAGGATAGCAGTGAATTCATCA 222
 35 sphe1y1s1a11e1y1s1e1u1y1s1a11e1g1y1G1n1A1s1p1e1s1e1r1u1l1e1n1 52
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 223 CTTCAAAGTCAAAATGACACACATCTCAAGAAACCAAGAAATCACTACT 272
 52 y1s1a11e1y1s1e1u1y1s1a11e1g1y1G1n1A1s1p1e1s1e1r1u1l1e1n1 68
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 273 GTCGAAACACAGGGGTTCCTCAATGATTCATCACTCAAGTTTCCTTTGAAGGT 322
 69 G1n1A1s1p1e1s1e1r1u1l1e1n1 83
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 323 CACGAAATGCTGATATATCACTACCCAAAGAACTGGCAATGAG 367

seq_name: /STD1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA43182

seq_documentation_block:

ID AAA43182 standard; cDNA; 335 BP.

AC AAA43182;

DT 21-AUG-2000 (first entry)

XX Xenopus secreted expressed sequence tag SEQ ID NO:1922.

XX Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;

KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytoskeletal; antibacterial; antifungal;
 KW antiviral; antidiabetic; antitumour; antiparkinsonian;
 KW anticancer; osteoprotective; neuroprotective; nootropic; antiproliferative;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.
 XX
 OS Xenopus sp.
 XX
 XX WO200021990-A1.
 XX
 PD 20-Apr-2000.
 XX
 XX 15-OCT-1999; 99WO-US24205.
 XX
 PF 15-OCT-1999; 99WO-US24205.
 XX
 PR 15-OCT-1998; 98US-0104435.
 XX
 PA (GEMV) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallee ER, Collins-Racie LA, Evans C;
 PI Metberg D, Treacy M;
 DR WPI; 2000-317937/27.
 XX
 XX Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 XX
 XX
 PS Claim 1; Page 559; 618pp; English.

AAA41261 to AAA43419 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytoskeletal; antibacterial; antifungal; antiviral; antidiabetic;
 CC antitumour; antiparkinsonian; antiproliferative; neuroprotective;
 CC nootropic; antiparkinsonian; antiproliferative; cerebroprotective;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiency, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
 CC in the exemplification of the present invention.

Sequence 335 BP; 117 A; 61 C; 86 G; 71 T; 0 other;

alignment_scores:

Quality: 32.00 Length: 32
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AAA43182 ..

Align seg 1/1 to: AAA43182 from: 1 to: 335

40 MetThrThrHisLeuLysLysLeuLysGluSerTyrCysGlnArgGlnGlu 56
 |||||
 153 ATGACACACGCAATCTCAAAAAGCTGAAAGAGTCATCTGTCAGAGACACAGG 202
 |||||
 56 yValProMetAsnSerLeuArgPheLeuPheGluGluGlnArgIle 71
 |||||
 203 CGTTCCAATGAATTCCTCAGGTTTGTGTTGAAGGCAAGAATC 248
 |||||

OM of: US-09-484-964-2 to: Issued_Patents_NA:* out_format : pfs
Date: Aug 15, 2001 6:40 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09484964/runat_14082001_111933_27708/app_query.fasta_1.158
-DB=Issued_Patents_NA -QPMT=fastap -SUFFIX=oli.rni -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-DEIOP=6.000 -EGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DEIOP=6.000 -DEIEXT=7.000 -START=1 -MATRIX=oligo
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=quality
-THR_MIN=20 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09484964 @CGNL_1_56 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLIPY -WAIT -THREADS=1

Search information block:

Query: US-09-484-964-2
Query Length: 101
Database: Issued_Patents_NA:*
Database sequences: 324599
Database length: 94655562
Search time (sec): 79.730000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence Strd Orig ZScore EScore Len i Documentation
/cgn2_6/ptodata/2/lna/5B_COMB.seq:US-09-213-768-1 + 83.00 1549.08 1.1e-78 1514 i
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-09-213-768-1

seq_documentation_block:

; Sequence 1, Application US/09213768
; Patent No. 5985664
; GENERAL INFORMATION:
; APPLICANT: Brenda F. Baker
; TITLE OF INVENTION: ANTISENSE MODULATION OF SENTRIN EXPRESSION
; FILE REFERENCE: RTS-0026
; CURRENT APPLICATION NUMBER: US/09/213,768
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (136)..(441)
US-09-213-768-1

alignment_scores:

Quality: 83.00 Length: 83
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x US-09-213-768-1 ..

Align seg 1/1 to: US-09-213-768-1 from: 1 to: 1514

19 GlyGluTyrIleLysLeuLysValIleGlyGlnAspSerGluIleH1 35
|||||
190 GGTGAATATATTAACTCAAGTCATTGGACGATGACAGTCA 239

35 sPheLysValLysMetThrThrHisLeuLysLysLeuYsgLusertYrc 52
|||||
240 CTTCAAAGTGAATAATGACACACATCTCAAGAACTCAAGATCATACT 289
52 ysgLinarGlnGlyValPrometasnSerLeuArgPheLeuPheGlnGly 68
|||||
290 GTCAAAGACAGGGGTTCATGAATTCACATGAGTTCTCTTGAGGGT 339
69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGlnGlnG 85
|||||
340 CAGAGAAATTCGATATATCATCTCCAAAAGAACTGGGATGAGGAGAGA 389
85 uasPValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThrVal 101
|||||
390 AGATGTGATTGAAGTTTATCAGGAACAACGGGGGTCTATTCAACAGTT 438

Thu Aug 16 08:48:52 2001

us-09-484-964-2.oli.rni

Page 2

OM of: US-09-484-964-2 to: EST:* out_format : pfs

Date: Aug 15, 2001 6:21 PM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

MODEL=frame+g2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09484964/runat_14082001_111932_27678/app_query.fasta_1.158
-DB=EST -OFAST=fastap -SUFFIX=collist -GAPOP=4.500 -GAPEXT=0.050
-MINMATCH=0.100 -LOCPCL=0.000 -LOEPEX=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000 -RGAPOP=6.000
-RGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=collist -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=quality -THR.MIN=20 -ALIGN=15
-MODE=LOCAL -OUTFMT=pfs -MINLEN=0 -MAXLEN=2000000000
-USER=US09484964 -CGN1_1_2867 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-484-964-2

Query length: 101

Database: EST*

Database sequences: 10228115

Database length: 431459454

Search time (sec): 1164.220000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
gb_est47:AA473590	+	83.00	1689.78	5.3e-85	321	AA473590 v63h04.r1 Beddington
gb_est4:AA227354	+	83.00	1688.94	5.0e-85	361	AA227354 z12211.r1 Stratagene
gb_est4:AA227355	+	83.00	1688.80	6.0e-85	368	AA227355 z12211.r1 Stratagene
gb_est81:BE982210	-	83.00	1688.70	6.1e-85	373	BE982210 UT-M-CGP-bf1-d-02-0-U
gb_est17:AT180895	+	83.00	1688.42	6.3e-85	388	AT180895 ub77ad5.r1 Soares_mamm
gb_est14:AA981271	+	83.00	1688.35	6.3e-85	391	AA981271 vx60f07.r1 Stratagene
gb_est1:AA421208	+	83.00	1688.16	6.5e-85	402	AA421208 z05h05.r1 Soares_test
gb_est11:AA726795	+	83.00	1688.07	6.6e-85	402	AA726795 v42910.r1 Barstead mc
gb_est1:AA450748	+	83.00	1688.00	6.6e-85	407	AA450748 v178d04.r1 Soares_mamm
gb_est88:BF512106	-	83.00	1687.85	6.8e-85	411	BF512106 UT-H-BMI-am1-9-09-0-UT
gb_est17:AT180896	-	83.00	1687.85	6.8e-85	420	AT180896 ub77ad5.r1 Soares_mamm
gb_est13:AA896203	+	83.00	1687.74	6.9e-85	426	AA896203 vx62d07.r1 Stratagene
gb_est15:AT047190	+	83.00	1687.69	6.9e-85	429	AT047190 vx62d04.r1 Soares_embd
gb_est195:BF978928	+	83.00	1687.68	6.9e-85	430	BF978928 602147602P1 NIH.MGC.62
gb_est10:AA682161	+	83.00	1687.66	6.9e-85	431	AA682161 v013e09.r1 Barstead mc
gb_est110:W24066	+	83.00	1687.63	7.0e-85	433	W24066 z082d02.r1 Soares_sensc
gb_est7:AA423388	+	83.00	1687.61	7.0e-85	434	AA423388 vx63b12.r1 Soares_mamm
gb_est13:AT164104	-	83.00	1687.53	7.1e-85	438	AT164104 ub77ad5.r1 Soares_mamm
gb_est16:BF333866	+	83.00	1687.53	7.1e-85	439	BF333866 SNEST420F08.Y1.CS1
gb_est10:AA636655	+	83.00	1687.48	7.1e-85	442	AA636655 v15f02.r1 Barstead mc
gb_est15:BE482568	+	83.00	1687.44	7.1e-85	444	BE482568 168404.BARC.5BOV.Bos.t
gb_est1:AA401864	+	83.00	1687.36	7.2e-85	449	AA401864 zve5b12.r1 Soares_totg
gb_est17:BE629336	+	83.00	1687.35	7.2e-85	450	BE629336 uc37cd2.r1 Soares_mamm
gb_gssl:AA046515	+	83.00	1687.35	7.2e-85	450	AA046515 AF046515 Mus.musculu
gb_est106:NA31695	+	83.00	1687.25	7.3e-85	456	NA31695 yx69g01.r1 Soares_melanc
gb_est1:AA105474	+	83.00	1687.22	7.4e-85	459	AA105474 mm32d10.r1 Stratagene
gb_est13:BE331886	+	83.00	1687.17	7.4e-85	461	BE331886 ut05c09.y1 Soares_thym
gb_est17:BE629336	+	83.00	1687.10	7.5e-85	466	BE629336 uc37d12.y1 Soares_mamm
gb_est17:BE629336	+	83.00	1687.02	7.5e-85	471	BE629336 168990.BARC.5BOV.Bos.t
gb_est17:BE197453	+	83.00	1686.87	7.7e-85	481	BE197453 ug74f06.y1 Soares_mamm
gb_est148:AW519530	+	83.00	1686.75	7.8e-85	489	AW519530 up33h12.y1 Soares_mous
gb_est1:AA030059	+	83.00	1686.72	7.8e-85	491	AA030059 m12h03.r1 Soares_mous
gb_est1:AA399812	+	83.00	1686.72	7.8e-85	491	AA399812 v07f112.r1 Beddington
gb_est15:AW323392	+	83.00	1686.65	7.9e-85	491	AW323392 uc08g05.y1 NCI.CGAP.Lu
gb_est17:AT1248769	-	83.00	1686.65	7.9e-85	496	AT1248769 qh77h07.y1 Soares_feta
gb_est14:AW239296	+	83.00	1686.65	7.9e-85	496	AW239296 qh77h07.y1 NCI.CGAP.Lu
gb_est15:AW989449	+	83.00	1686.65	7.9e-85	496	AW989449 ug14c01.y1 Soares_mamm
gb_est1:AW740709	+	83.00	1686.57	8.0e-85	501	AW740709 ur05b06.y1 NCI.CGAP.Mz
gb_est12:BE226598	+	83.00	1686.57	8.0e-85	501	BE226598 ia23c07.y1 Mouse.E10.5

seq_name: gb_est8:AA473590

seq_documentation_block:

LOCUS AA473590 321 bp mRNA EST 18-JUN-1997

DEFINITION v63h04.r1 Beddington mouse embryonic region Mus musculus cDNA clone IMAGE:866071 5' similar to TR:G927779 G927779 SUPPRESSOR OF

ACCESSION M122 MUTATIONS. ; mRNA sequence.

VERSION AA473590

KEYWORDS AA473590 1 GI:2201817

SOURCE EST.

ORGANISM Mus musculus

house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 321)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

PROJECT The WashU-HMI Mouse EST Project

UNPUBLISHED (1996)

CONTACT: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNC; contact the IMAGE Consortium (info@imgl.lnl.gov) for further information.

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 304.

Location/Qualifiers

1. 321

/organism="Mus musculus"

/strain="C57BL6 x DBA"

/db_xref="taxon:10090"

/clone="IMAGE:866071"

/clone_lib="Beddington mouse embryonic region"

/sex="pooled"

/tissue_type="embryo"

/dev_stage="7.5dpc"

/lab_host="DH12S"

/note="Organ: whole embryo; Vector: pcMV-SPOrt; Site:1; SalI: Site:2; NotI: Cloned unidirectionally. Primer: Oligo dT. Gastrulating embryos were collected at 7.5dpc from C57BL6 x DBA matings, excluding embryos that had developed head folds and all extraembryonic tissues. Average insert size: 1.3 kb (range: 0.5 - 3.0 kb). Referenced in Development 121, 2479-2489 (1995)."

BASE COUNT 118 a 51 c 75 g 77 t

ORIGIN

alignment_scores:

Quality: 83.00 Length: 83

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-484-964-2 x AA473590 ..

Align seg 1/1 to: AA473590 from: 1 to: 321


```

19 GlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGluIleH 35
   |||
56 GGAGAAATCAATTAACCAAGATTATGGACAGATAGCATGAGATAC 105
   |||
35 sphenylsValLysMetThrThrHisLeuLysLysLeuLysGluSerTyrC 52
   |||
106 TTTCAGAGTGAATAATGACACATCTCCAGAAACTCAAGAAATCATCTACT 155
   |||
52 ysglnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheGluGly 68
   |||
156 GTCAGAAAGACAGGAGTTCAGATTCAGTTCTCTCTTGAAGG 205
   |||
69 GlnArgIleAlaAspAsnHisThrProLysGlnLeuGlyMetGluGlu 85
   |||
206 CAGAGAAATTCCTGATTAATCATCTCCAGAAAGACTGGAGATGAGAGAA 255
   |||
85 uaspyAlIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101
   |||
256 AGATGTGATTGAAGTTTATTCAGAAACAGCGGGCTCATCTCAGCGGT 304
   |||
seq_name: gb_est4:AA227354

```

```

seq_documentation_block:
LOCUS AA227354 361 bp mRNA EST 24-FEB-1997
DEFINITION zrz2e12.r1 Striatogene NR2 neuronal precursor 937230 Homo sapiens
CDNA clone IMAGE:664172 5' similar to TR:G927779 G927779 SUPPRESSOR
OF MIF2 MUTATIONS. ;, mRNA sequence.
ACCESSION AA227354
VERSION AA227354.1 GI:1848909
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE 1 (bases 1 to 361)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
,M., Holtzman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevisan,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royally-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 282.
Location/Qualifiers
1. 361
/organism="Homo sapiens"
/db_xref="GDB:5426089"
/db_xref="taxon:9606"
/clone="IMAGE:664172"
/clone_lib="Striatogene NR2 neuronal precursor 937230"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
ECORT; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/c1.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGACGAG
3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTTCTT 3'"

```

```

BASE COUNT 121 a 68 c 92 g 80 t
ORIGIN

```

```

alignment_scores:
Quality: 83.00 Length: 83
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-484-964-2 x AA227354
..
Align seg 1/1 to: AA227354 from: 1 to: 361

```

```

19 GlyGluTyrIleLysLeuLysValIleGlyGlnAspSerSerGluIleH 35
   |||
108 GGTGAATTAATTAACCAAGATTCATGGACAGATAGCATGAGATTA 157
   |||
35 sphenylsValLysMetThrThrHisLeuLysLysLeuLysGluSerTyrC 52
   |||
158 CTCAAGTGAATAATGACACATCTCCAGAAACTCAAGAAATCATCTACT 207
   |||
52 ysglnArgGlnGlyValPrometAsnSerLeuArgPheLeuPheGluGly 68
   |||
208 GTCAGAAAGACAGGAGTTCAGATTCAGTTCTCTCTTGAAGG 257
   |||
69 GlnArgIleAlaAspAsnHisThrProLysGlnLeuGlyMetGluGlu 85
   |||
258 CAGAGAAATTCCTGATTAATCATCTCCAGAAAGACTGGAGATGAGAGAA 307
   |||
85 uaspyAlIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101
   |||
308 AGATGTGATTGAAGTTTATTCAGAAACAGCGGGCTCATCTCAGCGGT 356
   |||
seq_name: gb_est4:AA227355

```

```

seq_documentation_block:
LOCUS AA227355 368 bp mRNA EST 11-MAR-1998
DEFINITION zrz2e12.r1 Striatogene NR2 neuronal precursor 937230 Homo sapiens
CDNA clone IMAGE:664174 5' similar to TR:G927779 G927779 SUPPRESSOR
OF MIF2 MUTATIONS. ;, mRNA sequence.
ACCESSION AA227355
VERSION AA227355.1 GI:1848910
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE 1 (bases 1 to 368)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Getzel,G., Jost,S.,
Kizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royally-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1260 Std Error: 0.00
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 360.
Location/Qualifiers
1. 368
/organism="Homo sapiens"
/db_xref="GDB:5426092"
/db_xref="taxon:9606"
/clone="IMAGE:664174"
/clone_lib="Striatogene NR2 neuronal precursor 937230"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"

```

```

FEATURES
source

```



```

69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
|||||
264 CAGAGATTCCTGATATCATCTACCCGAAAGAACTGGCAATGAGGAGAGA 313
85 uaspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101
|||||
314 AGATGATTTGAAGTTATTCAGAACAAACGGGGGCTACTGACGGTT 362

seq_name: gb_est1:AA421208

seq_documentation_block:
LOCUS AA421208 402 bp mRNA EST 16-OCT-1997
DEFINITION zu05h05.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:731001
5' similar to TR:G927779 G927779 SUPPRESSOR OF MIF2 MUTATIONS. ;
mRNA sequence.
ACCESSION AA421208 GI:2100033
VERSION AA421208
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 402)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
T., Waterston, R., Wilson, R.
WashU-Merck EST Project 1997
JOURNAL
Unpublished (1997)
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 391.
Location/Qualifiers
1. 402
/organism="Homo sapiens"
/db_xref="GDB:5927801"
/db_xref="taxon:9606"
/clone_lib="IMAGE:731001"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAAGTGGAGCGCGCCATCTTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Col5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 134 a 74 c 105 g 89 t
ORIGIN

```

```

alignment_scores:
Quality: 83.00 Length: 83
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-484-964-2 x AA421208 ..
Align seg 1/1 to: AA421208 from: 1 to: 402

```

```

19 GlnGluTyrIleLysLeuLysValIleGlyGlnAspSerGluIleH1 35
|||||
147 GGTGATATATTAACATACTCAAGTACATGGACAGATFACGAGATTTCA 196
35 sPhelYsValLysMetThrThrHisLeuLysLysLeuYsgLusEryC 52
|||||
197 CTTCAAGATGAAATGACANACACATCTCAAGAACTCAAGAAATCATACT 246
52 ysgLInArgGlnGlyValPrometAsnSerLeuArgPheLeuPheLugly 68
|||||
247 GTCAAGACAGAGGTTCATCAATGAATTCATCAGGTTCCTTTGAGGCT 296
69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
|||||
297 CAGAGATTCCTGATATCATCTACCCGAAAGAACTGGCAATGAGGAGAGA 346
85 uaspValIleGluValTyrGlnGluGlnThrGlyGlyHisSerThrVal 101
|||||
347 AGATGATTTGAAGTTATTCAGAACAAACGGGGGCTACTGACGGTT 395

seq_name: gb_est1:AA726795

seq_documentation_block:
LOCUS AA726795 402 bp mRNA EST 02-JAN-1998
DEFINITION vu42q10.r1 Barstead mouse myotubes MRLB5 Mus musculus cDNA clone
IMAGE:1194114 5' similar to SW:SM33_HUMAN Q93068 UBIQUITIN-LIKE
PROTEIN SMT3C ; mRNA sequence.
ACCESSION AA726795
VERSION AA726795 GI:2744502
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 402)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisels, S., Kucaba, T., Lacy, M., Le, N., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HM Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HM Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:641210
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 385.
Location/Qualifiers
1. 402
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone_lib="IMAGE:1194114"
/clone_lib="Barstead mouse myotubes MRLB5"
/cell_line="C2C12"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAAGTGGAGCGCGCCATCTTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(PATTCGATCTTCG), digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing

```

BASE COUNT 136 a 72 c 104 g 90 t
 ORIGIN

alignment_scores:
 Quality: 83.00 Length: 83
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-484-964-2 x AA726795

Align seg 1/1 to: AA726795 from: 1 to: 402

19 GlyGluTyrIleLeuLeuValIleGlyGlnAspSerSergIuIleH1 35
 |||||
 139 GGAGATACATTAACTCAAGTTATTGGACAGATAGCAGTGAAGTACA 188
 |||||
 35 sPhelysValIysMetThrThriHisLeuLysLeuLysGluSerTyrC 52
 |||||
 189 TTTCAAGTGAATAATGACACACATCTCAAGAACTCAAGATCATACT 238
 |||||
 52 YsgIlnArgGlnGlyValPrometAsnSerLeuArpHeuNegIugly 68
 |||||
 239 GTCAAGACAGGAGTTCATCATCTCAAGGTTCTCTTGAAGGT 288
 |||||
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
 |||||
 289 CAGAGATTCGTGATATCATCTCCGAAGAAGCTGGAATGAGAGACAGA 338
 |||||
 85 uAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThrVal 101
 |||||
 339 AGATGTGATTGAAGTTATTCAGAACAAACGGGGGCTCACTGACGGGT 387

seq_name: gb_est7:AA450748

seq_documentation_block:

LOCUS AA450748 407 bp mRNA EST 04-JUN-1997
 DEFINITION V178604.1 Soares_mammary_gland_NBMWG Mus musculus cDNA clone
 IMAGE:849895 5' similar to TR:G927779 G927779 SUPPRESSOR OF MIF2
 MUTATIONS: ; mRNA sequence.

ACCESSION AA450748
 VERSION AA450748.1 GI:2164418

KEYWORDS EST.
 SOURCE house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 407)

REFERENCE
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucada, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellander, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE The MusH-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@wustl.edu
 This clone is available royalty-free through LIND; contact the
 IMAGE Consortium (info@image.lind.gov) for further information.
 MGI:502047

Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 387.

FEATURES

1. 407
 /organism="Mus musculus"
 /strain="C57BL/6J"

/db_xref="taxon:10090"
 /clone="IMAGE:849895"
 /clone_lib="Soares_mammary_gland_NBMWG"
 /sex="male"
 /tissue_type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: mammary gland; Vector: pT73d-Pac (Pharmacia)
) with a modified polylinker; Site 1: Not I; Site 2: Eco
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)
 primer [5',
 TGTTCACCAATCTGAAGTGGAGCCGCCGAATGCTTTTCTTTTCTTTTCTTTT
 T 3'] double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT73 vector.
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
 constructed and normalized by Bento Soares and M. Fatima
 Bonaldo."

BASE COUNT 134 a 72 c 104 g 97 t
 ORIGIN

alignment_scores:
 Quality: 83.00 Length: 83
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-484-964-2 x AA450748

Align seg 1/1 to: AA450748 from: 1 to: 407

19 GlyGluTyrIleLeuLeuValIleGlyGlnAspSerSergIuIleH1 35
 |||||
 137 GGAGATACATTAACTCAAGTTATTGGACAGATAGCAGTGAAGTACA 186
 |||||
 35 sPhelysValIysMetThrThriHisLeuLysLeuLysGluSerTyrC 52
 |||||
 187 TTTCAAGTGAATAATGACACACATCTCAAGAACTCAAGATCATACT 236
 |||||
 52 YsgIlnArgGlnGlyValPrometAsnSerLeuArpHeuNegIugly 68
 |||||
 237 GTCAAGACAGGAGTTCATCATCTCAAGGTTCTCTTGAAGGT 286
 |||||
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
 |||||
 287 CAGAGATTCGTGATATCATCTCCGAAGAAGCTGGAATGAGAGACAGA 336
 |||||
 85 uAspValIleGluValTyrGlnGlnGlnThrGlyGlyHisSerThrVal 101
 |||||
 337 AGATGTGATTGAAGTTATTCAGAACAAACGGGGGCTCACTGACGGGT 385

seq_name: gb_est88:BF512106

seq_documentation_block:

LOCUS BF512106 411 bp mRNA EST 07-DEC-2000
 DEFINITION UI-B-BW1-emi-9-09-0-UI.s1 NCI_CGAP_Sub7 Homo sapiens cDNA clone
 IMAGE:3070264 3', mRNA sequence.

ACCESSION BF512106
 VERSION BF512106.1 GI:11597318

KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 411)

REFERENCE
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaps-remail.nih.gov
 The sequence contained an oligo-dT track that was present in the

US-09-484-964-2 x A1180896

Align seg 1/1 to: A1180896 from: 1 to: 420

19 GlyGluTrpIleLeuLysLeuValIleGlyGlnAspSerSerGluIleH1 35
 143 GGAGAAATGCTAACTCAAGTATTGGACAGATAGCATGATATCA 192
 35 sPhelysValLysMetThrHisLeuLysLysLeuGluSerTyrC 52
 193 TTTCAAGTGAATAATGACACATCTCAAGAAATCAAGATCATACT 242
 52 ysglnarqngllyvalprometasenleuargpheleupheglugly 68
 243 GTCAAGAGACAGGAGTCCATATTCATCTCAGGTTTCTTTGAAGGT 292
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
 293 CAGAGAAATGCTGATATCATCTCCGAAAGAACTGGAATGAGAGAGA 342
 85 uaspvalilegluvaltyrtingluginthrlyglyhisserthrval 101
 343 AGATGTGATTGAAGTTTATCAGAACAAACGGGGGTCTCAGCGGTT 391

seq_name: gb_est13:AA896203

seq_documentation_block:

LOCUS AA896203 426 bp mRNA EST 06-APR-1998
 DEFINITION vx2d07.r1 Striatogene mouse macrophage (#937306) Mus musculus cDNA

Clone IMAGE:1279789 5' similar to SW:SM33_HUMAN Q93068
 UBIQUITIN-LIKE PROTEIN SMT3C ; mRNA sequence.

ACCESSION AA896203
 VERSION AA896203.1 GI:3032596

KEYWORDS EST.
 SOURCE house mouse.

ORGANISM

Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 426)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu

Mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:671589

Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 407.

location/Qualifiers

FEATURES

SOURCE

1. 426
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:1279789"
 /clone_lib="Striatogene mouse macrophage (#937306)"
 /tissue_type="macrophage"
 /dev_stage="MEH1-3 cell line"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ blood; Vector: pBluescript SK-; Site: 1:
 EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. WEH1-3 cell line. Average insert size: 1.5 kb;
 Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGCGACGAC
 3' ~3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'."
 139 a 80 c 114 g 93 t

ORIGIN

alignment_scores:
 Quality: 83.00 Length: 83
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-484-964-2 x AA896203

Align seg 1/1 to: AA896203 from: 1 to: 426

19 GlyGluTrpIleLeuLysLeuValIleGlyGlnAspSerSerGluIleH1 35
 170 GGAGAAATGCTAACTCAAGTATTGGACAGATAGCATGATATCA 219
 35 sPhelysValLysMetThrHisLeuLysLysLeuGluSerTyrC 52
 220 TTTCAAGTGAATAATGACACATCTCAAGAAATCAAGATCATACT 269
 52 ysglnarqngllyvalprometasenleuargpheleupheglugly 68
 270 GTCAAGAGACAGGAGTCCATATTCATCTCAGGTTTCTTTGAAGGT 319
 69 GlnArgIleAlaAspAsnHisThrProLysGluLeuGlyMetGluGlu 85
 320 CAGAGAAATGCTGATATCATCTCCGAAAGAACTGGAATGAGAGAGA 369
 85 uaspvalilegluvaltyrtingluginthrlyglyhisserthrval 101
 370 AGATGTGATTGAAGTTTATCAGAACAAACGGGGGTCTCAGCGGTT 418

seq_name: gb_est15:A1047190

seq_documentation_block:

LOCUS A1047190 429 bp mRNA EST 08-JUN-1998
 DEFINITION uh52d04.r1 Soares embryonic stem cell NMES Mus musculus cDNA clone

IMAGE:1749991 5' similar to SW:SM33_HUMAN Q93068 UBIQUITIN-LIKE
 PROTEIN SMT3C ; mRNA sequence.

ACCESSION A1047190
 VERSION A1047190.1 GI:3295477

KEYWORDS EST.
 SOURCE house mouse.

ORGANISM

Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 429)

REFERENCE

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu

Mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:961803

Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 407.

location/Qualifiers

FEATURES

SOURCE

1. 429
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:1749991"
 /clone_lib="Soares_embryonic_stem_cell_NMES"

```

/cell_type="embryonic stem cell"
/lab_host="DH10B"
/note="Organ: bone marrow; Vector: p7T73D-Pac (pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'
TCTTACCAATCTGAAGTGGAGCGCCGATGTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p7T73 vector. Library
went through two rounds of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      140 a      76 c      109 g      104 t
ORIGIN
alignment_scores:
  quality:      83.00      length:      83
  ratio:        1.000      gaps:        0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
  US-09-484-964-2 x AT047190
Align seg 1/1 to: AT047190 from: 1 to: 429
19 GYGLUTYRILEYSLYSLYVAIIIEGLYGLNAPSERSERGIIIEH 35
|||||
154 GGAGAAATACATTAACCTCAAGATTGACAGAGATGACAGTGAATCA 203
35 SPHELYSVALISMETHRTHHISLEULYSLEULYSGLUSERTYRC 52
|||||
204 TTTCAAAGTGAATGACACACATCTCAAGAAATCAAAATCATCT 253
52 YSGLNARGINGLYVALPROMETASNSERLEUARGPHELEUPHEGLUG 68
|||||
254 GTCAAAAGCAGGGAGTCCAAATGATTCAGTCAAGTTCTCTTGAAGGT 303
69 GINARGILEALASPASNHSIHPROLYSGIULEUGLYMETGLUGI 85
|||||
304 CAGAGAAATGCTGATATCATCTCCGAAAGAACTGGGAATGGAGAGA 353
85 UASPYAIIIEGLUVALTYRGINGUGINTHRGILYGLYHISSETRHVAL 101
|||||
354 AGATGATTTGAAGTTATTCAGAGAACAAAGGGGGTCACTCCAGCGTT 402
seq_name: gb_est195:BF978928
seq_documentation_block:
  LOCUS      BF978928      430 bp      mRNA      EST      22-JAN-2001
  DEFINITION  602147602F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4306489 5',
  ACCESSION  BF978928
  VERSION    BF978928.1 GI:12346143
  KEYWORDS
  SOURCE      human.
  ORGANISM    Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 430)
  NIH-MGC http://mhc.nci.nih.gov/
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: ATCC/DC/DT/DTP
  CDNA Library Preparation: CLONTECH Laboratories, Inc.
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/MLN at:
  http://image.llnl.gov
  Plate: L1CM1175 row: c column: 02

```

```

FEATURES
  source
    Location/Qualifiers
      1..430
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:4306489"
      /clone_lib="NIH_MGC_62"
      /tissue_type="melanotic melanoma, high MDR"
      /lab_host="DH10B (T1 phage-resistant)"
      /note="Organ: skin; Vector: pBNR-LIB (Clontech); Site_1:
      SfiI (ggccatcgcc); Site_2: SfiI (ggccatcgcc);
      Double-stranded cDNA was prepared from cell line RNA. 5'
      and 3' adaptors were used in cloning as follows: 5'
      adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor
      sequence: 5'-ATTCTAGAGCGCCGCGGCGGACATG-dt(30)BN-3'
      (where B = A, C, or G and N = A, C, G, or T). Average
      insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
      contained inserts by PCR. This library was enriched for
      full-length clones and was constructed by Clontech
      Laboratories (Palo Alto, CA)."
BASE COUNT      134 a      80 c      119 g      97 t
ORIGIN
alignment_scores:
  quality:      83.00      length:      83
  ratio:        1.000      gaps:        0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
  US-09-484-964-2 x BF978928
Align seg 1/1 to: BF978928 from: 1 to: 430
19 GYGLUTYRILEYSLYSLYVAIIIEGLYGLNAPSERSERGIIIEH 35
|||||
171 GGTGATATATTAATCAAGATGACAGAGATGACAGATTCATCA 220
35 SPHELYSVALISMETHRTHHISLEULYSLEULYSGLUSERTYRC 52
|||||
221 CTTCAAGAGTGAATGACACACATCTCAAGAACTCAAGAAATCATCT 270
52 YSGLNARGINGLYVALPROMETASNSERLEUARGPHELEUPHEGLUG 68
|||||
271 GTCAAAAGCAGGGAGTCCAAATGATTCAGTCAAGTTCTCTTGAAGGT 320
69 GINARGILEALASPASNHSIHPROLYSGIULEUGLYMETGLUGI 85
|||||
321 CAGAGAAATGCTGATATCATCTCCAAAGAACTGGGAATGGAGAGA 370
85 UASPYAIIIEGLUVALTYRGINGUGINTHRGILYGLYHISSETRHVAL 101
|||||
371 AGATGATTTGAAGTTATTCAGAGAACAAAGGGGGTCACTCCAGCGTT 419
seq_name: gb_est10:AA682161
seq_documentation_block:
  LOCUS      AA682161      431 bp      mRNA      EST      05-DEC-1997
  DEFINITION  v013609.r1 Barsstead mouse myotubes MRLRB5 Mus musculus cDNA clone
  IMAGE:1180552 5' similar to TR:G927779 G927779 SUPPRESSOR OF MIF2
  MUTATIONS.: , mRNA sequence.
  ACCESSION  AA682161
  VERSION    AA682161.1 GI:2663193
  KEYWORDS
  SOURCE      house mouse.
  ORGANISM    Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 431)
  Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
  Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
  Schellenberg, K., Steptoe, M., Tan, F., Underwood, R., Moore, B.,
  Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

```